

日 本 国 特 許 庁
JAPAN PATENT OFFICE

Jc978 U.S. PTO
10/083354
02/27/02

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2001年11月12日

出 願 番 号
Application Number:

特願2001-345525

[ST.10/C]:

[JP2001-345525]

出 願 人
Applicant(s):

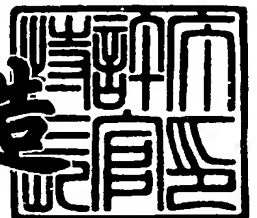
株式会社日立製作所

U. S. Appln. Filed 2-27-02
Inventor: K. Mogi et al
Mattingly Stanger & Malor
Docket ASA-1071

2002年 1月11日

特 許 庁 長 官
Commissioner,
Japan Patent Office

及 川 耕 造



出証番号 出証特2001-3114289

【書類名】 特許願

【整理番号】 K01011381A

【あて先】 特許庁長官

【国際特許分類】 G06F 17/30

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

【氏名】 茂木 和彦

【発明者】

【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社日立製作所 システム開発研究所内

【氏名】 大枝 高

【発明者】

【住所又は居所】 千葉県松戸市二十世紀が丘丸山町 1 7

【氏名】 喜連川 優

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 100075096

【弁理士】

【氏名又は名称】 作田 康夫

【手数料の表示】

【予納台帳番号】 013088

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データベース管理システムの静的な情報を取得する手段を有する記憶装置

【特許請求の範囲】

【請求項 1】

データベース管理システムが稼動している計算機との接続手段を有し、

前記データベース管理システムにおけるスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベース管理システムが管理するデータベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報を取得する情報取得手段を有することを特徴とする記憶装置。

【請求項 2】

前記接続手段を用いて複数の前記データベース管理システムが稼動している計算機と接続することを特徴とする請求項 1 に記載の記憶装置。

【請求項 3】

前記情報取得手段が複数の前記データベース管理システムが管理するデータベースに関する情報を取得することを特徴とする請求項 1 に記載の記憶装置。

【請求項 4】

前記情報取得手段は前記接続手段を用いて情報を取得することを特徴とする請求項 1 に記載の記憶装置。

【請求項 5】

前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムから取得することを特徴とする請求項 1 に記載の記憶装置。

【請求項 6】

前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムとは異なる少なくとも 1 つのプログラムを通して取得することを特徴とする請求項 1 に記載の記憶装置。

【請求項 7】

前記記憶装置は少なくとも1つ以上のデータを記憶する物理記憶手段を有し、

前記計算機が前記記憶装置をアクセスする際に利用する論理的な格納位置を前記記憶装置内で実際にデータを記憶する物理記憶手段の記憶位置へ変換する論理-物理位置変換手段を有し、

前記論理位置に対応するデータの前記物理記憶手段における記憶位置を変更するデータの配置変更手段を有し、

前記情報取得手段により取得した情報を利用する前記論理位置に対応するデータの物理記憶位置の変更案を作成する配置変更案作成手段を有することを特徴とする請求項1に記載の記憶装置。

【請求項8】

前記配置変更手段を用いて前記配置変更案作成手段により作成されたデータの配置の変更を行なう自動データ再配置手段を有することを特徴とする請求項7に記載の記憶装置。

【請求項9】

前記配置変更案作成手段が、前記情報取得手段により取得した情報に基づいて、前記データベース管理システムが前記データベースデータをシーケンシャルにアクセスする際のアクセス場所とアクセス順を判断し、前記シーケンシャルにアクセスされるデータベースデータを前記物理記憶手段上で前記連続アクセス順の前後関係を崩さずに連続した領域に配置することを特徴とする請求項7に記載の記憶装置。

【請求項10】

前記情報取得手段が取得する情報は、前記データベース管理システムが前記スキーマにより定義される同一のデータ構造に属する前記データベースデータをアクセスする際の並列度に関する情報を含み、

前記配置変更案作成手段が、前記取得情報を基に、前記スキーマにより定義される同一のデータ構造に属する前記データベースデータを複数の前記物理記憶手段に配置することを特徴とする請求項7に記載の記憶装置。

【請求項11】

前記配置変更案作成手段が、前記取得情報に基づいて、同時にアクセスされる

可能性が高い前記データベースデータの組を検出し、それらを異なる前記物理記憶手段に配置することを特徴とする請求項 7 に記載の記憶装置。

【請求項 1 2】

前記配置変更案作成手段が、表データと前記表データに対する索引データを抽出し、前記索引が木構造のデータ構造の場合にそれらを前記同時にアクセスされる可能性が高いデータベースデータの組として登録することを特徴とする請求項 1 1 に記載の記憶装置。

【請求項 1 3】

前記データベースに関する情報に、前記データベース管理システムにおける処理の実行履歴に関する情報を含むことを特徴とする請求項 1 1 に記載の記憶装置。

【請求項 1 4】

前記物理記憶手段の稼動情報を取得する物理記憶手段稼動情報取得手段を有し、
前記配置変更案作成手段が前記物理記憶手段稼動情報取得手段により取得した情報を利用することを特徴とする請求項 1 1 に記載の記憶装置。

【請求項 1 5】

前記配置変更案作成手段が、前記データベース管理システムにおけるデータの更新処理時に記録するログデータとその他の前記データベースデータを前記同時にアクセスされる可能性が高いデータベースデータの組として登録することを特徴とする請求項 1 1 に記載の記憶装置。

【請求項 1 6】

キャッシュメモリとデータを記憶する物理記憶手段を有し、
前記情報取得手段により取得した情報を利用して前記キャッシュメモリを制御するキャッシュメモリ制御手段を有し、
前記情報取得手段が取得する情報として、前記データベース管理システムが前記計算機上のメモリ上に有するキャッシュの前記データベースデータに対する制御方法とそのキャッシュデータ量に関する情報である計算機キャッシュデータ情報を含むことを特徴とする請求項 1 に記載の記憶装置。

【請求項 1 7】

前記キャッシュメモリ制御手段が、前記スキーマにより定義されるデータ構造に関して前記計算機上メモリにキャッシュされているデータ量である計算機データ構造キャッシュデータ量と前記データ構造の実データのデータサイズを比較し、前記比較結果を用いて前記データ構造の実データ記憶位置に記憶されているデータのキャッシュ優先度を決定することを特徴とする請求項 1 6 に記載の記憶装置。

【請求項 1 8】

前記キャッシュメモリ制御手段が、前記計算機データ構造キャッシュデータ量と前記記憶装置における前記データ構造に属するデータに対して前記キャッシュメモリの利用可能量を比較し、前記比較結果を用いて前記データ構造の実データ記憶位置に記憶されているデータのキャッシュ優先度を決定することを特徴とする請求項 1 6 に記載の記憶装置。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、データベース管理システムに関する。

【0 0 0 2】

【従来の技術】

現在、データベース（DB）を基盤とする多くのアプリケーションが存在し、DBに関する一連の処理・管理を行うソフトウェアであるデータベース管理システム（DBMS）は極めて重要なものとなっている。特に、DBMSの処理性能はDBを利用するアプリケーションの性能も決定するため、DBMSの処理性能の向上は極めて重要である。

【0 0 0 3】

DBの特徴の1つは、多大な量のデータを扱うことである。そのため、DBMSの実行環境の多くにおいては、DBMSが実行される計算機に対して大容量の記憶装置を接続し、記憶装置上にDBのデータを記憶する。そのため、DBに関する処理を行う場合に、記憶装置に対してアクセスが発生し、記憶装置における

データアクセス性能がDBMSの性能を大きく左右する。そのため、DBMSが稼動するシステムにおいて、記憶装置の性能最適化が極めて重要である。

【0004】

文献“Oracle 8i パフォーマンスのための設計およびチューニング、リリース 8.1”（日本オラクル社、部品番号 J00921-01）の第 20 章（文献 1）においては、RDBMS である Oracle 8i における I/O のチューニングについて述べられている。その中で、RDBMS の内部動作のチューニングと共に、データの配置のチューニングに関連するものとして、ログファイルは他のデータファイルから分離した物理記憶装置に記憶すること、ストライプ化されたディスクにデータを記憶することによる負荷分散が効果があること、表のデータとそれに対応する索引データは異なる物理記憶装置に記憶すると効果があること、RDBMS とは関係ないデータを異なる物理記憶装置に記憶することが述べられている。

【0005】

米国特許 6035306（文献 2）においては、DBMS-ファイルシステム-ボリュームマネージャー記憶装置間のマッピングを考慮した性能解析ツールに関する技術を開示している。この性能解析ツールは、各レイヤにおけるオブジェクトの稼動状況を画面に表示する。このときに上記のマッピングを考慮し、その各オブジェクトに対応する他レイヤのオブジェクトの稼動状況を示す画面を容易に表示する機能を提供する。また、ボリュームマネージャレイヤのオブジェクトに関して、負荷が高い記憶装置群に記憶されているオブジェクトのうち、2 番目に負荷が高いオブジェクトを、もっとも負荷が低い記憶装置群に移動するオブジェクト再配置案を作成する機能を有している。

【0006】

特開平 9-274544 号公報（文献 3）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理記憶装置上に配置する記憶装置において、前記論理的記憶装置の物理記憶装置への配置を動的に変更することにより記憶装置のアクセス性能を向上する技術について開示している。アクセス頻度が高い物理記憶装置に記憶されているデータの一部を前記の配置

動的変更機能を用いて他の物理記憶装置に移動することにより、特定の物理記憶装置のアクセス頻度が高くなるようにし、これにより記憶装置を全体としてみたときの性能を向上させる。また、配置動的変更機能による高性能化処理の自動実行方法についても開示している。

【 0 0 0 7 】

特開 2 0 0 1 - 6 7 1 8 7 号公報（文献 4）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理記憶装置上に配置し、前記論理的記憶装置の物理記憶装置への配置を動的に変更する機能を有する記憶装置において、論理的記憶装置の物理記憶装置への配置の変更案を作成する際に物理記憶装置を属性の異なるグループに分割し、それを考慮した配置変更案を作成し、その配置変更案に従って自動的に論理的記憶装置の配置を変更する技術について開示している。配置変更案作成時に、物理記憶装置を属性毎にグループ化し、論理的記憶装置の配置先として、それが有している特徴にあった属性を保持している物理記憶装置のグループに属する物理記憶装置を選択する配置変更案を作成することによりより良好なものを作成する。

【 0 0 0 8 】

米国特許 5 3 1 7 7 2 7（文献 5）においては、DBMS の処理の一部あるいは全部を記憶装置側で実施することにより DBMS の高速化する技術について開示している。記憶装置側で DBMS 処理を行うため、記憶装置においてデータアクセス特性を把握することが可能であり、データアクセス特性と記憶装置の構成を考慮することにより無駄な物理記憶媒体に対するアクセスを削減や必要なデータの先読みを実施することができ、結果として DBMS の性能を向上させることができる。

【 0 0 0 9 】

論文“高機能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価”（向井他著、第 1 1 回データ工学ワークショップ（DEWS 2 0 0 0））論文集講演番号 3 B - 3、2 0 0 0 年 7 月発行 CD - ROM、主催：電子情報通信学会データ工学研究専門委員会）（文献 6）では、記憶装置の高機能化による DBMS の性能向上について論じている。具体的には、記憶装置に対してアプ

リケーションレベルの知識としてリレーショナルデータベース管理システム（RDBMS）における問い合わせ処理の実行時のデータのアクセスプランを与えた場合の効果について述べている。更に、その確認のために、トレースデータを用いたホスト側から発行する先読み指示により前記技術を模した擬似実験を実施している。

【 0 0 1 0 】

【発明が解決しようとする課題】

従来の技術には以下のような問題が存在する。

【 0 0 1 1 】

文献 1 で述べられているものは、管理者がデータの配置を決定する際に考慮すべき項目である。現在、1 つの記憶装置内に多数の物理記憶装置を有し、多数の計算機により共有されるものが存在する。この記憶装置においては、多くの場合、ホストが認識する論理的記憶装置を実際にデータを記憶する物理記憶装置の適当な領域に割り当てることが行われる。このような記憶装置を利用する場合、人間がすべてを把握することは困難であり、このような記憶装置を含む計算機システム側に何かしらのサポート機能が存在しなければ文献 1 に述べられている問題点を把握することすら困難となる。また、問題点を把握することができたとしても、計算機システム側にデータの移動機能が存在しない場合には、記憶装置上のデータの再配置のためにデータのバックアップとリストアが必要となり、その処理に多大な労力を必要とする。

【 0 0 1 2 】

文献 2 で述べられている技術では、ボリュームマネージャレイヤにおけるオブジェクトの稼動状況による配置最適化案を作成する機能を実現しているが、記憶装置から更に高いアクセス性能を得ようとする場合には DBMS レイヤにおけるデータの特徴を考慮して配置を決定する必要があるがその点の解決方法に関しては何も述べていない。

【 0 0 1 3 】

文献 3、文献 4 で述べられている技術に関しては、記憶装置におけるデータ記憶位置の最適化に関する技術である。これらの技術においては記憶装置を利用す

るアプリケーションが利用するデータに関する特徴としては、アクセス頻度とシーケンシャルアクセス性程度しか考慮していないため、アプリケーションから見た場合に必ずしも最適な配置が実現できるわけではない。例えば、DBMSにおける表データとそれに対応する索引データのような同時にアクセスされるデータを同一の物理記憶装置に配置する可能性がある。このとき、その物理記憶装置においてアクセス競合が発生し、記憶装置のアクセス性能が低下する可能性がある。

また、文献1から文献4で述べられている技術に関しては、記憶装置におけるキャッシュメモリの利用については特に考慮されていない。

【0014】

文献5、文献6で述べられている技術に関しては、データの記録位置の最適化に関して特に考慮をしていない。そのため、特定の物理記憶装置の負荷が高くなっているために記憶装置の性能が低下し、DBMSの性能が悪化している状況の解決手段として有効なものではない。

【0015】

本発明の第一の目的は、DBMSが管理するデータを保持する記憶装置において、DBMSの処理の特徴を考慮することによりDBMSに対してより好ましい性能特性を持つ記憶装置を実現することである。この記憶装置を利用することにより、既存のDBMSを利用したDBシステムに対しても、DBMSの性能を向上させることができる。

【0016】

本発明の第二の目的は、記憶装置の性能最適化機能を提供することにより記憶装置の性能に関する管理コストを削減することである。これにより、この記憶装置を用いたDBシステムのシステム管理コストを削減することができる。

【0017】

【課題を解決するための手段】

DBMSに関する情報を記憶装置が取得し、その情報と記憶装置内の物理記憶装置の特性と更に利用可能であれば記憶装置を利用する他のアプリケーションのアクセス頻度に関する情報を考慮する記憶装置の性能最適化処理を記憶装置上で

実施する。

【 0 0 1 8 】

記憶装置がDBMSの特性を考慮してDBMSに良好な性能を得るための手段の第一として、ホストが認識する論理的記憶装置を物理記憶装置の適当な領域に割り当ててデータを記憶する記憶装置における、論理的記憶装置のデータ記憶位置の最適化が存在する。例えば、データ更新時に必ず書き込みが実行される更新ログを、他のデータと異なる物理記憶装置に配置して相互干渉しないようにすることによりDBMSに対して良好な性能特性を得ることができる。

【 0 0 1 9 】

また、同時にアクセスされる可能性が極めて高い表データとそれに対応する索引データを異なる物理記憶装置に配置することによりDBMSに対して良好な性能特性を得ることができる。更に、DBMSに関する情報を利用して、データがシーケンシャルにアクセスされる場合のアクセス順序を予測し、その構造を保持するように物理記憶装置に記憶することによりシーケンシャルアクセス性能を向上可能である。現在、論理的記憶装置の記憶位置を動的に変更する技術は存在し、これを利用することによりデータの最適配置を実現できる。

【 0 0 2 0 】

記憶装置がDBMSの特性を考慮してDBMSに良好な性能を得るための手段の第二として、DBMSにおけるホスト上のキャッシュ動作を考慮したキャッシュメモリ制御が存在する。DBMSにおいては利用頻度の高いデータをホストのメモリ上にキャッシュするが、全てのデータがホストメモリ上に乗ってしまうようなデータに対しては、記憶装置上のキャッシュに保持してもあまり効果はない。

【 0 0 2 1 】

また、多くのDBMSにおいては、ホスト上のキャッシュの破棄データの選択にLRUアルゴリズムを用いている。ホスト上にキャッシュ可能なデータ量と比べてある一定量以下のデータしか記憶装置上のキャッシュに保持できない場合は、読み出しアクセスにより記憶装置上のキャッシュ上に保持された後にキャッシュに乗っている間に再利用される可能性は低く、そのようなデータを記憶装置上

のキャッシュに保持することの効果は小さい。このようなデータをキャッシュから優先的に破棄するような制御を記憶装置上で行うことにより、キャッシュ効果の高いものをより多量に記憶装置のキャッシュメモリ上に保持できるようになり、記憶装置のアクセス性能が向上する。

【 0 0 2 2 】

【発明の実施の形態】

以下、本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

＜第一の実施の形態＞

本実施の形態では、DBMSが実行される計算機と記憶装置が接続された計算機システムにおいて、記憶装置がDBMSに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報を取得し、それらを用いて記憶装置の動作を改善する。記憶装置において、記憶装置内部でデータの記憶位置を動的に変更する機能を有し、取得した情報をもとに好適なデータ再配置案を作成し、データの記憶位置の動的変更機能を用いて、作成したデータ再配置案に従ったデータ配置を実現し、アクセス性能を改善する。また、取得情報をもとにしたデータキャッシュの制御を行いより良いアクセス性能特性が得られるようにする。

【 0 0 2 3 】

図1は、本発明の第一の実施の形態における計算機システムの構成図である。本実施の形態における計算機システムは、DBホスト80a、80b、ホスト情報設定サーバ82、記憶装置10から構成される。DBホスト80a、80b、ホスト情報設定サーバ82、記憶装置10はそれぞれが保有するネットワークインターフェイス78を通してネットワーク79に接続されている。また、DBホスト80a、80b、記憶装置10はそれぞれが保有するI/Oバスインターフェイス70からI/Oバス71を介してI/Oバススイッチ72に接続され、これらを通して記憶装置10とDBホスト80a、80b間のデータ転送を行う。

【 0 0 2 4 】

本実施の形態においては、記憶装置10とDBホスト80a、80b間のデータ転送を行うI/Oバス71とネットワーク79を異なるものとしているが、例

例えば i S C S I のような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されており、本実施の形態においてもこの技術を利用してもよい。このとき、記憶装置 1 0 と DB ホスト 8 0 a, 8 0 b において I / O パスインターフェイス 7 0 が省かれ、計算機システム内から I / O パス 7 1 と I / O パススイッチ 7 2 が省かれる構成となる。

【 0 0 2 5 】

記憶装置 1 0 は、記憶領域を提供するもので、その記憶領域は記憶領域管理単位であるボリュームを用いて外部に提供し、ボリューム内の部分領域に対するアクセスや管理はブロックを単位として実行する。記憶装置 1 0 は、ネットワークインターフェイス 7 8、I / O パスインターフェイス 7 0、記憶装置制御装置 1 2、ディスクコントローラ 1 6、物理記憶装置 1 8 から構成され、ネットワークインターフェイス 7 8、I / O パスインターフェイス 7 0、記憶装置制御装置 1 2、ディスクコントローラ 1 6 はそれぞれ内部バス 2 0 により接続され、ディスクコントローラ 1 6 と物理記憶装置 1 8 は物理記憶装置バス 2 2 により接続される。記憶装置制御装置 1 2 は、CPU 2 4 とメモリ 2 6 を有する。

【 0 0 2 6 】

メモリ 2 6 上には、記憶装置におけるキャッシュメモリとして利用するデータキャッシュ 2 8 が割り当てられ、記憶装置を制御するためのプログラムである記憶装置制御プログラム 4 0 が記憶される。また、メモリ 2 6 上には、物理記憶装置 1 8 の稼動情報である物理記憶装置稼動情報 3 2、データキャッシュ 2 8 の管理情報であるデータキャッシュ管理情報 3 4、DB ホスト 8 0 a, 8 0 b で実行されている DBMS 1 1 0 a, 1 1 0 b に関する情報である DBMS データ情報 3 6、記憶装置 1 0 が提供するボリュームを物理的に記憶する物理記憶装置 1 8 上の記憶位置の管理情報であるボリューム物理記憶位置管理情報 3 8 を保持する。

【 0 0 2 7 】

図中の記憶装置 1 0 は、複数の物理記憶装置 1 8 を有し、1 つのボリュームに属するデータを複数の物理記憶装置 1 8 に分散配置することが可能である。また、データが記憶される物理記憶装置 1 8 上の位置を動的に変更する機能を有する。

。記憶装置制御プログラム40は、ディスクコントローラ16の制御を行うディスクコントローラ制御部42、データキャッシュ28の管理を行うキャッシュ制御部44、記憶装置10が提供するボリュームを物理的に記憶する物理記憶装置18上の記憶位置の管理やデータが記憶される物理記憶装置18上の位置を動的に変更する機能に関する処理を行う物理記憶位置管理・最適化部46、I/Oバスインターフェイス70の制御を行うI/Oバスインターフェイス制御部48、ネットワークインターフェイス78の制御を行うネットワークインターフェイス制御部50を含む。

【0028】

DBホスト80a、80b、ホスト情報設定サーバ82においては、それぞれCPU84、ネットワークインターフェイス78、メモリ88を有し、メモリ88上にオペレーティングシステム(OS)100が記憶・実行されている。

【0029】

DBホスト80a、80bはI/Oバスインターフェイス70を有し、記憶装置10が提供するボリュームに対してアクセスを実行する。OS100内にファイルシステム104と1つ以上のボリュームからホストが利用する論理的なボリュームである論理ボリュームを作成するボリュームマネージャ102と、ファイルシステム104やボリュームマネージャ102により、OS100によりアプリケーションに対して提供されるファイルや論理ローボリュームに記憶されたデータの記録位置等を管理するマッピング情報106を有する。OS100が認識するボリュームやボリュームマネージャ102により提供される論理ボリュームに対して、アプリケーションがそれらのボリュームをファイルと等価なインターフェイスでアクセスするための機構であるロードバース機構をOS100が有していても良い。

【0030】

図中の構成ではボリュームマネージャ102が存在しているが、本実施の形態においてはボリュームマネージャ102における論理ボリュームの構成を変更することはないので、ボリュームマネージャ102が存在せずにファイルシステムが記憶装置10により提供されるボリュームを利用する構成に対しても本実施の

形態を当てはめることができる。

【0031】

DBホスト80a, 80bのそれぞれのメモリ88上ではDBMS110a, 110bが記憶・実行され、実行履歴情報122が記憶されている。DBMS110a, 110bは内部にスキーマ情報114を有している。図中では、DBMS110a, 110bが1台のホストに1つのみ動作しているが、後述するように、DBMS110a, 110b毎の識別子を用いて管理を行うため、1台のホストにDBMSが複数動作していても本実施の形態にあてはめることができる。

【0032】

DBホスト80a上ではDBMS情報取得・通信プログラム118が動作している。一方、DBホスト80b上ではDBMS情報取得・通信プログラム118が提供する機能をDBMS110b中のDBMS情報収集・通信部116が提供する。

【0033】

ホスト情報設定サーバ82のメモリ88上ではホスト情報設定プログラム130が記憶・実行される。

【0034】

図2はDBホスト80a, 80bのOS100内に記憶されているマッピング情報106を示す。マッピング情報106中には、ボリュームローデバース情報520、ファイル記憶位置情報530と論理ボリューム構成情報540が含まれる。

【0035】

ボリュームローデバース情報520中にはOS100においてローデバースを指定するための識別子であるローデバースパス名521とそのローデバースによりアクセスされる記憶装置10が提供するボリュームあるいは論理ボリュームの識別子であるローデバースボリューム名522の組が含まれる。

【0036】

ファイル記憶位置情報530中には、OS100においてファイルを指定するための識別子であるファイルパス名531とそのファイル中のデータ位置を指定

するブロック番号であるファイルブロック番号532とそれに対応するデータが記憶されている記憶装置10が提供するボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名533とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号534の組が含まれる。

【0037】

論理ボリューム構成情報540中にはボリュームマネージャ102により提供される論理ボリュームの識別子である論理ボリューム名541とその論理ボリューム上のデータの位置を示す論理ボリューム論理ブロック番号542とその論理ブロックが記憶されているボリュームの識別子であるボリューム名501とボリューム上の記憶位置であるボリューム論理ブロック番号512の組が含まれる。マッピング情報106を取得するには、OS100が提供している管理コマンドの実行や情報提供機構の利用、場合によっては参照可能な管理データの直接解析等を行う必要がある。

【0038】

図3はDBMS110a, 110b内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報114を示す。スキーマ情報114には、表のデータ構造や制約条件等の定義情報を保持する表定義情報551、索引のデータ構造や対象である表等の定義情報を保持する索引定義情報552、利用するログに関する情報であるログ情報553、利用する一時表領域に関する情報である一時表領域情報554、管理しているデータのデータ記憶位置の管理情報であるデータ記憶位置情報555、キャッシュの構成に関する情報であるキャッシュ構成情報556とデータをアクセスする際の並列度に関する情報である最大アクセス並列度情報557を含む。

【0039】

データ記憶位置情報555中には、表、索引、ログ、一時表領域等のデータ構造の識別子であるデータ構造名561とそのデータを記憶するファイルまたはローデバイスの識別子であるデータファイルパス名562とその中の記憶位置であるファイルブロック番号563の組が含まれる。キャッシュ構成情報556はDBMS110a, 110bが三種類のキャッシュ管理のグループを定義し、そのグ

ループに対してキャッシュを割り当てている場合を示す。

【0040】

キャッシュ構成情報556中には、グループ名565とグループ中のデータ構造のデータをホスト上にキャッシュする際の最大データサイズであるキャッシュサイズ566とそのグループに所属するデータ構造の識別子の所属データ構造名567の組が含まれる。最大アクセス並列度情報557には、データ構造名561とそのデータ構造にアクセスする際の一般的な場合の最大並列度に関する情報である最大アクセス並列度569の組が含まれる。

【0041】

スキーマ情報114を外部から取得するには、管理ビューとして外部に公開されているものをSQL等のデータ検索言語を用いて取得したり、または、専用の機構を用いて取得したりすることができる。

【0042】

図4はDBホスト80a、80bのメモリ88上に記憶されている実行履歴情報122を示す。実行履歴情報122中には、DBMS110a、110bで実行されたクエリ570の履歴が記憶されている。この情報は、DBMS110a、110bで作成する。またはDBMSのフロントエンドプログラムがこの情報を作成する。この場合には、DBMSフロントエンドプログラムが存在する計算機に実行履歴情報122が記憶されることになる。

【0043】

図5は記憶装置10内に保持されているボリューム物理記憶位置管理情報38を示す。ボリューム物理記憶位置管理情報38中には、データの論理アドレス-物理記憶装置18における記憶位置のマッピングを管理するボリューム物理記憶位置メイン情報510と記憶装置10内でのボリュームに属するデータの物理記憶位置の変更処理の管理情報であるボリュームデータ移動管理情報511が含まれる。

【0044】

ボリューム物理記憶位置メイン情報510中には、ボリューム名501とそのボリューム上のデータ記憶位置であるボリューム論理ブロック番号512とその

論理ブロックが記憶されている物理記憶装置18の識別子である物理記憶装置名502と物理記憶装置18上の記憶位置である物理ブロック番号514の組のデータが含まれる。ここで、ボリューム名501が“Empty”であるエントリ515は特殊なエントリであり、このエントリには記憶装置10内の物理記憶装置18の領域のうち、ボリュームに割り当てられていない領域を示し、この領域に対してデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【0045】

ボリュームデータ移動管理情報511はボリューム名501と、そのボリューム内の記憶位置を変更するデータ範囲を示す移動論理ブロック番号782と、そのデータが新規に記憶される物理記憶装置18の識別子とその記憶領域を示す移動先物理記憶装置名783と移動先物理ブロック番号784、現在のデータコピー元を示すコピーポインタ786とデータの再コピーの必要性を管理する差分管理情報785の組が含まれる。

【0046】

差分管理情報785とコピーポインタ786を用いたデータの記憶位置変更処理の概略を以下に示す。差分管理情報785はある一定量の領域毎にデータコピーが必要である「1」または不必要「0」を示すデータを保持する。データの記憶位置変更処理開始時に全ての差分管理情報785のエントリを1にセットし、コピーポインタ786を移動元の先頭にセットする。コピーポインタ786にしたがって差分管理情報785に1がセットされている領域を順次移動先にデータをコピーし、コピーポインタ786を更新していく。差分管理情報785で管理される領域をコピーする直前に、その対応するエントリを0にセットする。

【0047】

データコピー中に移動領域内のデータに対する更新が行われた場合、それに対応する差分管理情報785のエントリを1にセットする。一度全領域のコピーが完了した段階で差分管理情報785内のエントリが全て0になったかを確認し、全て0であればボリューム物理記憶位置メイン情報510を更新してデータの記憶位置変更処理は完了する。1のエントリが残っている場合には、再度それに対

応する領域をコピーする処理を前記手順で繰り返す。

【 0 0 4 8 】

なお、データ記憶位置の動的変更機能の実現方法は他の方式を用いても良い。
この場合には、ボリューム物理記憶位置管理情報 3 8 中にはボリュームデータ移動管理情報 5 1 1 ではなく他のデータ記憶位置の動的変更機能のための管理情報が含まれることになる。

【 0 0 4 9 】

図 6 に記憶装置 1 0 内に保持されている物理記憶装置稼働情報 3 2 を示す。物理記憶装置稼働情報 3 2 中には、記憶装置 1 0 が提供するボリュームの識別子であるボリューム名 5 0 1 とそのボリューム名 5 0 1 を持つボリュームのデータを保持する物理記憶装置 1 8 の識別子である物理記憶装置名 5 0 2、そしてボリューム名 5 0 1 を持つボリュームが物理記憶装置名 5 0 2 を持つ物理記憶装置 1 8 に記憶しているデータをアクセスするための稼働時間のある時刻からの累積値である累積稼働時間 5 0 3、稼働率 5 9 4 計算のために前回利用した累積稼働時間 5 0 3 の値である旧累積稼働時間 5 9 3 とある一定時間内の動作時間の割合を示す稼働率 5 9 4 の組と、稼働率 5 9 4 計算のために前回累積稼働時間を取得した時刻である前回累積稼働時間取得時刻 5 9 5 を含む。

【 0 0 5 0 】

ディスクコントローラ制御部 4 2 はディスクコントローラ 1 6 を利用して物理記憶装置 1 8 へのデータアクセスする際の開始時刻と終了時刻を取得し、そのアクセスデータがどのボリュームに対するものかを判断して開始時刻と終了時刻の差分を稼働時間として対応するボリューム名 5 0 1 と物理記憶装置名 5 0 2 を持つデータの組の累積稼働時間 5 0 3 に加算する。

【 0 0 5 1 】

物理記憶位置管理・最適化部 4 6 は一定間隔で以下の処理を行う。累積稼働時間 5 0 3 と旧累積稼働時間 5 9 3、前回累積稼働時間取得時刻 5 9 5 と現データ取得時刻を用いて前回累積稼働時間取得時刻 5 9 5 と現データ取得時刻間の稼働率 5 9 4 を計算・記憶する。その後、取得した累積稼働時間 5 0 3 を旧累積稼働時間 5 9 3 に、現データ取得時刻を前回累積稼働時間取得時刻 5 9 5 に記憶する

【0052】

図7に記憶装置10内に保持されているDBMSデータ情報36を示す。DBMSデータ情報36中には、DBMSスキーマ情報711、データ構造物理記憶位置情報712、DBMS実行履歴情報714、DBMSデータ構造キャッシュ効果情報715を含む。

【0053】

DBMSデータ情報36中に含まれるデータは、DBホスト80a、80b上に存在するデータを利用する必要があるものが含まれる。記憶装置10は記憶装置10の外部に存在する情報をホスト情報設定サーバ82で動作するホスト情報設定プログラム130を利用して取得する。ホスト情報設定プログラム130はネットワーク79を通し、DBホスト80a上で実行され、マッピング情報106等必要となる情報の収集処理を実施するDBMS情報取得・通信プログラム118や、DBホスト80b上で実行されているDBMS110b中のDBMS情報取得・通信プログラム118と等価な機能を実現するDBMS情報収集・通信部116を利用して必要な情報を収集する。

【0054】

ホスト情報設定プログラム130は情報取得後、必要ならば記憶装置10に情報を設定するためのデータの加工を行い、ネットワーク79を通して記憶装置10に転送する。記憶装置10においては、ネットワークインターフェイス制御部50が必要な情報が送られてきたことを確認し、物理記憶位置管理・最適化部46に渡し、必要な加工を行った後にその情報をDBMSデータ情報36中の適切な場所に記憶する。

【0055】

ホスト情報設定プログラム130は任意のDBホスト80a、80b上で実行されていてもよい。あるいは、物理記憶位置管理・最適化部46がホスト情報設定プログラム130の情報収集機能を有してもよい。これらの場合は、DBホスト80a、80bから情報を転送する際にI/Oパス71を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリューム

を記憶装置 1 0 は DB ホスト 8 0 a, 8 0 b に提供し、そのボリュームに対する書き込みがあった場合に I/O パスインターフェイス制御部 7 0 は情報の転送があったと判断し、その情報を物理記憶位置管理・最適化部 4 6 に渡し、必要な加工を行った後にその情報を DBMS データ情報 3 6 中の適切な場所に記憶する、等の方式を利用する。

【 0 0 5 6 】

情報の収集処理に関しては、記憶装置 1 0 が必要になったときに外部にデータ転送要求を出す方法と、データの変更があるたびに外部から記憶装置 1 0 に変更されたデータを送る方法の 2 種類ともに利用することができる。

【 0 0 5 7 】

図 8 に DBMS データ情報 3 6 中に含まれる DBMS スキーマ情報 7 1 1 を示す。DBMS スキーマ情報 7 1 1 は、DBMS データ構造情報 6 2 1、DBMS データ記憶位置情報 6 2 2、DBMS パーティション化表・索引情報 6 2 3、DBMS 索引定義情報 6 2 4、DBMS キャッシュ構成情報 6 2 5、DBMS ホスト情報 6 2 6 を含む。

【 0 0 5 8 】

DBMS データ構造情報 6 2 1 は DBMS 1 1 0 a, 1 1 0 b で定義されているデータ構造に関する情報で、DBMS 1 1 0 a, 1 1 0 b の識別子である DBMS 名 6 3 1、DBMS 1 1 0 a, 1 1 0 b 内の表・索引・ログ・一時表領域等のデータ構造の識別子であるデータ構造名 5 6 1、データ構造の種別を表すデータ構造種別 6 4 0、データ記憶位置情報から求めることができるデータ構造が利用する総データ量を示すデータ構造データ量 6 4 1、そのデータ構造をアクセスする際の最大並列度に関する情報である最大アクセス並列度 5 6 9 の組を保持する。このとき、データ構造によっては最大アクセス並列度 5 6 9 の値を持たない。

【 0 0 5 9 】

DBMS データ記憶位置情報 6 2 2 は DBMS 名 6 3 1 とその DBMS におけるデータ記憶位置管理情報 5 5 5 であるデータ記憶位置管理情報 6 3 8 の組を保持する。

【0060】

DBMSパーティション化表・索引情報623は、1つの表や索引をある属性値により幾つかのグループに分割したデータ構造を管理する情報で、パーティション化されたデータ構造が所属するDBMS110a, 110bの識別子であるDBMS名631と分割化される前のデータ構造の識別子であるパーティション元データ構造名643と分割後のデータ構造の識別子であるデータ構造名561とその分割条件を保持するパーティション化方法644の組を保持する。今後、パーティション化されたデータ構造に関しては、特に断らない限り単純にデータ構造と呼ぶ場合にはパーティション化後のものを指すものとする。

【0061】

DBMS索引定義情報624には、DBMS名631、索引の識別子である索引名635、その索引のデータ形式を示す索引タイプ636、その索引がどの表のどの属性に対するものかを示す対応表情報637の組を保持する。

【0062】

DBMSキャッシュ構成情報625は、DBMS110a, 110bのキャッシュに関する情報であり、DBMS名631とDBMS110a, 110bにおけるキャッシュ構成情報556の組を保持する。

【0063】

DBMSホスト情報626は、DBMS名631を持つDBMS110a, 110bがどのホスト上で実行されているかを管理するもので、DBMS名631とDBMS実行ホストの識別子であるホスト名651の組を保持する。DBMSスキーマ情報711中のDBMSホスト情報626以外は、DBMS110a, 110bが管理しているスキーマ情報114の中から必要な情報を取得して作成する。DBMSホスト情報626はシステム構成情報で管理者が設定するものである。

【0064】

図9にDBMSデータ情報36中に含まれるデータ構造物理記憶位置情報712を示す。データ構造物理記憶位置情報712はDBMS110a, 110bに含まれるデータ構造が記憶装置10内でどの物理記憶装置18のどの領域に記憶

されるかを管理するもので、データ構造を特定するDBMS名631とデータ構造名561、その外部からのアクセス領域を示すボリューム名501とボリューム論理ブロック番号512、その物理記憶装置18上の記憶位置を示す物理記憶装置名502と物理ブロック番号514の組を保持する。この情報は、DBMSデータ記憶位置情報622とマッピング情報106を記憶装置10の外部から取得し、さらにボリューム物理記憶位置メイン情報510を参照して、対応する部分を組み合わせることにより作成する。

【0065】

DBMS110a, 110b毎にシーケンシャルアクセスの方法が定まっている。データ構造物理記憶位置情報712を作成する際に、DBMS名631とデータ構造名561により特定されるデータ構造毎に、シーケンシャルアクセス時のアクセス順を保持するようにソートしたデータを作成する。ここでは、対象とするDBMS110a, 110bの種類を絞り、あらかじめデータ構造物理記憶位置情報712を作成するプログラムがDBMS110a, 110bにおけるシーケンシャルアクセス方法を把握し、シーケンシャルアクセス順でソートされたデータを作成する。

【0066】

本実施の形態のDBMS110a, 110bにおけるシーケンシャルアクセス方法は以下の方法に従うものとする。あるデータ構造のデータをシーケンシャルアクセスする場合に、データ構造が記憶されているデータファイル名562とファイルブロック番号563を昇順にソートしその順序でアクセスを実行する。その他にシーケンシャルアクセス方法の決定方法としては、データファイルを管理する内部通番とファイルブロック番号563の組を昇順にソートした順番にアクセスする方法等が存在し、それらを利用したシーケンシャルアクセス方法の判断を実施してもよい。

【0067】

図10にDBMSデータ情報36中に含まれるクエリ実行同時アクセスデータ構造カウント情報714を示す。これは、実行履歴情報122をもとに同時にアクセスされるデータ構造の組と実行履歴中に何回その組を同時にアクセスするク

エリが実行されたかを示すデータで、DBMS名631、同時にアクセスされる可能性のあるデータ構造のデータ構造名561の組を示すデータ構造名A701とデータ構造名B702、そして、DBMS実行履歴122の解析によりそのデータ構造の組がアクセスされたと判断された回数であるカウント値703の組で表される。この組はカウント値703の値でソートしておく。

【0068】

クエリ実行時同時アクセスデータカウント情報714はDBMS実行履歴122から作成する。最初にクエリ実行時同時アクセスデータカウント情報714のエントリを全消去する。DBMS110a, 110bにおいて定型処理が行われる場合には、まず、その型により分類し、その型の処理が何回実行されたかを確認する。続いてDBMS110a, 110bから型毎のクエリ実行プランを取得する。そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。

【0069】

そして、クエリ実行時同時アクセスデータカウント情報714中のDBMS名631・データ構造名A701・データ構造名B702を参照し、既に対応するデータ構造の組が存在している場合には先に求めたその型の処理回数をカウント値703に加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリを追加してカウント値703を先に求めたその型の処理回数にセットする。

【0070】

DBMS110a, 110bにおいて非定型処理が行われる場合には、1つ1つの実行されたクエリに関してクエリ実行プランを取得し、そのクエリ実行プランにより示される処理手順から同時にアクセスされるデータ構造の組を判別する。そして、クエリ実行時同時アクセスデータカウント情報714中のDBMS名631・データ構造名A701・データ構造名B702を参照し、既に対応するデータ構造の組が存在している場合にはカウント値703に1を加算する。既に対応するデータ構造の組が存在していない場合には、新たにエントリを追加してカウント値703に1をセットする。

【 0 0 7 1 】

クエリ実行プランから同時にアクセスされる可能性があるデータ構造の判別は以下のように行う。まず、木構造の索引に対するアクセスが実施される場合には、その木構造索引データと、その索引が対象とする表データが同時にアクセスされると判断する。また、データの更新処理や挿入処理が行われる場合には、ログとその他のデータが同時にアクセスされると判断する。

【 0 0 7 2 】

以下はDBMS 1 1 0 a, 1 1 0 bの特性に依存するが、例えば、クエリ実行プラン作成時にネストループジョイン処理を多段に渡り実行する計画を作成し、それらの多段に渡る処理を同時に実行するRDBMSが存在する。このRDBMSを利用する場合にはその多段に渡るネストループジョイン処理で利用する表データとその表に対する木構造の索引データは同時にアクセスされると判断できる。

【 0 0 7 3 】

このように、クエリ実行計画による同時アクセスデータの判断に関しては、DBMS 1 1 0 a, 1 1 0 bの処理特性を把握して判断する必要があるが、ここでは、対象とするDBMS 1 1 0 a, 1 1 0 bの種類を絞り、クエリ実行時同時アクセスデータカウント情報 7 1 4 を作成するプログラムがDBMS 1 1 0 a, 1 1 0 b特有の同時アクセスデータ構造の組を把握できる機能を有することを仮定する。

【 0 0 7 4 】

実行履歴情報 1 2 2 からクエリ実行時同時アクセスデータカウント情報 7 1 4 を作成する処理は、記憶装置 1 0 の内部、外部どちらで実行してもよい。記憶装置 1 0 でクエリ実行時同時アクセスデータカウント情報 7 1 4 を作成する場合には、記憶装置 1 0 がネットワーク 7 9 を通してDBホスト 8 0 a, 8 0 b、あるいは、実行履歴情報 1 2 2 がDBMSフロントエンドプログラムが実行される計算機上に記憶される場合にはその計算機に対して実行履歴情報 1 2 2 を記憶装置 1 0 に転送する要求を出し、その情報をネットワーク 7 9 を通して受け取る。

【 0 0 7 5 】

その後、前述のクエリ実行時同時アクセスデータカウント情報 7 1 4 作成処理を実施する。記憶装置 1 0 の外部で作成する場合は、例えば、ホスト情報設定サーバ 8 2 が DB ホスト 8 0 a, 8 0 b、あるいは DBMS フロントエンドプログラムが実行される計算機から実行履歴情報 1 2 2 を取得し、クエリ実行時同時アクセスデータカウント情報 7 1 4 作成処理を実施する。その後、ネットワーク 7 9 を通して作成されたクエリ実行時同時アクセスデータカウント情報 7 1 4 を記憶装置 1 0 に転送し、それを DBMS データ情報 3 6 中に記憶する。

【 0 0 7 6 】

なお、本実施の形態においては、常に実行履歴情報 1 2 2 が作成される必要はない。クエリ実行時同時アクセスデータカウント情報 7 1 4 作成時に実行履歴情報 1 2 2 が存在しない DBMS 1 1 0 a, 1 1 0 b が利用するデータ構造に関してはそれらを見捨ててデータを作成する。また、クエリ実行時同時アクセスデータカウント情報 7 1 4 は存在しなくてもよい。

【 0 0 7 7 】

図 1 1 に DBMS データ情報 3 6 に含まれる DBMS データ構造キャッシュ効果情報 7 1 5 を示す。DBMS データ構造キャッシュ効果情報 7 1 5 は記憶装置 1 0 においてデータ構造をデータキャッシュに保持しておくことに効果があるかどうかを判断した結果を保持するもので、データ構造を特定する DBMS 名 6 3 1 とデータ構造名 5 6 1、そのデータ構造がデータキャッシュに保持する効果があるかどうかの判断結果を示すキャッシュ効果情報 7 3 3 を保持する。キャッシュ効果情報 7 3 3 の値は、管理者が指定する、もしくは、以下に述べる手順に従って求めるものである。

【 0 0 7 8 】

図 1 2 に記憶装置 1 0 において指定されたデータ構造のデータをデータキャッシュに保持する効果があるかどうかの判断する処理のフローを示す。判断基準は 2 種類有り、1 つは「指定データ構造のデータ量に比べてホストキャッシュ量が十分に存在するために利用頻度が高いデータの読出しアクセスが実行されないか」で、もう 1 つは「記憶装置 1 0 のデータキャッシュ量がホストキャッシュ量に比べて小さく、記憶装置 1 0 のデータキャッシュ量で効果がある利用頻度のデー

タはホストキャッシュに載ってしまい、記憶装置 1 0 から読出されるデータを記憶装置 1 0 でキャッシュしても効果が低い」ことである。

【 0 0 7 9 】

ステップ 2 8 0 1 で処理を開始する。ステップ 2 8 0 2 で指定データ構造と同じキャッシュ管理のグループに属するデータ構造のデータ量の総和を DBMS キャッシュ構成情報 6 2 5 と DBMS データ構造情報 6 2 1 を参照して求める。

【 0 0 8 0 】

ステップ 2 8 0 3 で指定データ構造と同じキャッシュ管理のグループにおけるそのグループの単位データ量あたりのホストにおける平均キャッシュ量を前記のグループのデータ総量と DBMS キャッシュ構成情報 6 2 5 中のキャッシュサイズ 5 6 6 から求め、その値をあらかじめ定められたキャッシュ効果判断閾値と比較する。その値が閾値以上の場合にはステップ 2 8 0 7 に進み、閾値未満の場合にはステップ 2 8 0 4 に進む。この閾値としては概ね 0. 7 程度の値を用いる。

【 0 0 8 1 】

ステップ 2 8 0 4 では記憶装置 1 0 における単位容量あたりの平均キャッシュデータ量を求める。この値は、記憶装置のデータキャッシュ 2 8 の総容量と外部に提供するボリュームの総容量から求めることができ、これらの値はボリューム物理記憶位置管理情報 3 8 やデータキャッシュ管理情報 3 4 を参照することにより求めることができる。

【 0 0 8 2 】

ステップ 2 8 0 5 では、前述の指定データ構造が属するキャッシュ管理のグループにおける単位データ量あたりのホストにおける平均キャッシュ量に対する記憶装置 1 0 における平均キャッシュ量の比率を求め、その値がキャッシュ効果判断閾値未満の場合はステップ 2 8 0 7 に進み、閾値以上の場合にはステップ 2 8 0 6 に進む。この閾値としては概ね 0. 7 程度の値を用いる。

【 0 0 8 3 】

ステップ 2 8 0 6 では記憶装置 1 0 においてキャッシュする効果があると判定し、ステップ 2 8 0 8 に進みキャッシュ効果判定処理を終了する。

【 0 0 8 4 】

ステップ2807では記憶装置10においてキャッシュする効果がないと判定し、ステップ2808に進みキャッシュ効果判定処理を終了する。

【0085】

記憶装置10は、データキャッシュをある一定サイズの領域であるセグメントと呼ぶ管理単位を用いて管理する。図13に記憶装置10内に保持されているデータキャッシュ管理情報34を示す。データキャッシュ管理情報34中には、データキャッシュ28のセグメントの状態を示すキャッシュセグメント情報720とキャッシュセグメントの再利用対象選定に利用するキャッシュセグメント利用管理情報740を含む。

【0086】

キャッシュセグメント情報720中には、セグメントの識別子であるセグメントID721と、そのセグメントに記憶されているデータ領域を示すボリューム名511とボリューム論理ブロック番号512、そして、セグメントの状態を示すステータス情報722、後述するセグメントの再利用選定管理に利用するリストの情報であるリスト情報723を含む。

【0087】

ステータス情報722が示すセグメントの状態としては、物理記憶装置18上にセグメント内のデータと同じデータが記憶されている“ノーマル”、セグメント内にのみ最新のデータが存在する“ダーティ”、セグメント内に有効なデータが存在しない“インバリッド”が存在する。リスト情報723には、現在そのセグメントが属するリストの識別子と、そのリストのリンク情報が記憶される。図中では、リストは双方向リンクリストであるとしている。

【0088】

キャッシュセグメント利用管理情報740中には、キャッシュセグメントの再利用対象選定に利用する3種類の管理リストである第1LRUリスト、第2LRUリスト、再利用LRUリストの管理情報として、第1LRUリスト情報741、第2LRUリスト情報742、再利用LRUリスト情報743が記憶される。

【0089】

第1LRUリスト情報741、第2LRUリスト情報742、再利用LRUリ

スト情報 7 4 3 は、それぞれリストの先頭である M R U セグメント I D、最後尾である L R U セグメント I D、そのリストに属するセグメント数を記憶する。この 3 種類の管理リストはホストからのアクセス要求の処理にかかわるもので、アクセス要求処理の説明時に同時に行う。

【 0 0 9 0 】

ホストからのデータアクセス要求があったときの処理を説明する。

【 0 0 9 1 】

図 1 4 に記憶装置 1 0 がホストからデータの読出し要求を受け取ったときの処理フローを示す。ステップ 2 9 0 1 で、I / O パスインターフェイス 7 0 はホストからのデータ読出し要求を受け、I / O パスインターフェイス制御部 4 8 がその要求を認識する。

【 0 0 9 2 】

ステップ 2 9 0 2 でキャッシュ制御部 4 4 は読出し要求があったデータがデータキャッシュ 2 8 上に存在するかデータキャッシュ管理情報 3 4 を参照して確認する。存在する場合にはステップ 2 9 0 5 に進み、存在しない場合にはステップ 2 9 0 3 に進む。

【 0 0 9 3 】

ステップ 2 9 0 3 で、キャッシュ制御部 4 4 は読出し要求があったデータを保持するキャッシュ領域を確保する。データを保持するキャッシュセグメントとして、ステータス情報がノーマルのもののうち、再利用 L R U リストの L R U (L e a s t R e c e n t l y U s e d : 最も昔に使われた) 側に存在するものを必要数取得し、再利用 L R U リストから削除する。そして、再利用 L R U リスト情報 7 4 3 をそれに合わせて更新する。また、キャッシュセグメント情報 7 2 0 中のボリューム名 5 1 1 とボリューム論理ブロック番号を記憶するデータのものに変更し、ステータス情報 7 2 2 をインバリッドに設定する。

【 0 0 9 4 】

ステップ 2 9 0 4 でディスクコントローラ制御部 4 2 は読出し要求があったデータを物理記憶装置 1 8 から読み出す処理を実施し、その完了を待つ。読出し完了後、キャッシュセグメント情報 7 2 0 中の対応するステータス情報 7 2 2 をノ

ーマルに設定し、ステップ 2 9 0 6 に進む。

【 0 0 9 5 】

ステップ 2 9 0 5 でキャッシュ制御部 4 4 はデータ読出し要求のあったデータを保持するセグメントをその管理のためにリンクされている管理リストから削除する。

【 0 0 9 6 】

ステップ 2 9 0 6 で I / O パスインターフェイス管理部 4 8 はデータ読出し要求のあったデータをセグメントから I / O パスインターフェイス 7 0 を利用してホストに転送し、ホストとの処理を完了する。

【 0 0 9 7 】

ステップ 2 9 0 7 でキャッシュ制御部 4 4 はアクセス先のデータの内容に従い、データ読出し要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を行う。この処理の詳細は後述する。

【 0 0 9 8 】

ステップ 2 9 0 8 でホストからの読出し要求を受けとった時の処理を終了する。

【 0 0 9 9 】

図 1 5 に記憶装置 1 0 がホストからデータの書き込み要求を受け取ったときの処理フローを示す。ステップ 2 9 3 1 で、I / O パスインターフェイス 7 0 はホストからのデータ書き込み要求を受け、I / O パスインターフェイス制御部 4 8 がその要求を認識する。

【 0 1 0 0 】

ステップ 2 9 3 2 でキャッシュ制御部 4 4 は読出し要求があったデータを保持するセグメントがデータキャッシュ 2 8 上に存在するかデータキャッシュ管理情報 3 4 を参照して確認する。存在する場合にはステップ 2 9 3 4 に進み、存在しない場合にはステップ 2 9 3 3 に進む。

【 0 1 0 1 】

ステップ 2 9 3 3 で、キャッシュ制御部 4 4 は書き込み要求があったデータを保持するキャッシュ領域を確保する。データを保持するキャッシュセグメントと

して、ステータス情報がノーマルのもののうち、再利用LRUリストのLRU側に存在するものを必要数取得し、再利用LRUリストから削除する。そして、再利用LRUリスト情報743をそれに合わせて更新する。また、キャッシュセグメント情報720中のボリューム名511とボリューム論理ブロック番号を記憶するデータのものに變更し、ステータス情報722をインバリッドに設定する。

【0102】

ステップ2934でキャッシュ制御部44はデータ書込み要求のあったデータを保持するセグメントをその管理のためにリンクされている管理リストから削除する。

【0103】

ステップ2935でI/Oパスインターフェイス管理部48はデータ書込み要求のあったデータをキャッシュセグメントに書込み、キャッシュセグメント情報720中の対応するステータス情報722をダーティに設定し、ホストとの処理を完了する。

【0104】

ステップ2936でキャッシュ制御部44はアクセス先のデータの内容に従い、データ書込み要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を行う。この処理の詳細は後述する。

【0105】

ステップ2937でホストからの書込み要求を受けとった時の処理を終了する。

【0106】

図16にキャッシュ制御部44が実行するアクセス先のデータの内容に従い、アクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理のフローを示す。この処理において、記憶装置10におけるキャッシュ効果がないと判断されるデータを保持するキャッシュセグメントを管理リスト中の再利用されやすい場所に繋ぐことによりキャッシュ効果がないと判断されるものがデータキャッシュ28上に載っている時間を短くし、他のデータのキャッシュ効果を高めることを行う。

【 0 1 0 7 】

ステップ 2 9 6 1 においてアクセス先のデータの内容に従い、アクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理を開始する。

【 0 1 0 8 】

ステップ 2 9 6 2 において、アクセス先データのキャッシュ効果の確認を行う。データ構造物理記憶位置情報 7 1 2 を参照してアクセス先データが属する DBMS 1 1 0 a, 1 1 0 b とそのデータ構造の識別子である DBMS 名 6 3 1 とデータ構造名 5 6 1 を求める。データ構造物理記憶位置情報 7 1 2 に対応部分がない場合にはキャッシュ効果があると判断する。

【 0 1 0 9 】

続いて、DBMS データ構造キャッシュ効果情報 7 1 5 を参照し、既に求めた DBMS 名 6 3 1 とデータ構造名 5 6 1 に対応するキャッシュ効果情報 7 3 3 を参照し、アクセス先データにキャッシュ効果があるかないかを求める。なお、キャッシュ効果情報 7 3 3 中に対応するエントリがない場合、キャッシュ効果があると判断する。キャッシュ効果があると判断された場合にはステップ 2 9 6 3 に進み、ないと判断された場合にはステップ 2 9 6 6 に進む。

【 0 1 1 0 】

ステップ 2 9 6 3 でアクセス先データを保持するキャッシュセグメントを第 1 LRU リストの MRU (Most Recently Used : 最も最近使われた) 側にリンクし、それにあわせて第 1 LRU リスト情報 7 4 1 を更新する。

【 0 1 1 1 】

ステップ 2 9 6 4 で第 1 LRU リストにリンクされているセグメント数を第 1 LRU リスト情報 7 4 1 を参照して確認し、その値が事前に定めてある閾値を超えているか確認する。そのセグメント数が閾値未満の場合にはステップ 2 9 7 0 に進み処理を完了する。閾値以上の場合にはステップ 2 9 6 5 に進む。

【 0 1 1 2 】

ステップ 2 9 6 5 で第 1 LRU リストのセグメント数が閾値未満になるように第 1 LRU の最も LRU 側に存在するセグメントを第 2 LRU リストの MRU 側にリンクし直す処理を行い、それに合わせて第 1 LRU リスト情報 7 4 1 と第 2

L R U リスト情報 7 4 2 を更新し、ステップ 2 9 6 7 に進む。

【 0 1 1 3 】

ステップ 2 9 6 6 でアクセス先データを保持するキャッシュセグメントを第 2 L R U リストの M R U 側にリンクし、それに合わせて第 2 L R U リスト情報 7 4 2 を更新し、ステップ 2 9 6 7 に進む。

【 0 1 1 4 】

ステップ 2 9 6 7 で第 2 L R U リストにリンクされているセグメント数を第 2 L R U リスト情報 7 4 2 を参照して確認し、その値が事前に定めてある閾値を超えているか確認する。そのセグメント数が閾値未満の場合にはステップ 2 9 7 0 に進み処理を完了する。閾値以上の場合にはステップ 2 9 6 8 に進む。

【 0 1 1 5 】

ステップ 2 9 6 8 で第 2 L R U リストのセグメント数が閾値未満になるように第 2 L R U の最も L R U 側に存在するセグメントを再利用 L R U リストの M R U 側にリンクし直す処理を行い、それに合わせて第 2 L R U リスト情報 7 4 2 と再利用 L R U リスト情報 7 4 3 を更新する。

【 0 1 1 6 】

ステップ 2 9 6 9 で、ステップ 2 9 6 8 において第 2 L R U リストから再利用 L R U リストにリンクしなおされたセグメントに関して、キャッシュセグメント情報 7 2 0 中のステータス情報 7 2 2 を参照して、その値がダーティであるもののデータを物理記憶装置 1 8 に書き出す処理をディスクコントローラ制御部 4 2 に対して要求し、その完了を待つ。書き出し処理完了後、キャッシュセグメント情報 7 2 0 中の対応するステータス情報 7 2 2 をノーマルに変更し、ステップ 2 9 7 0 に進む。

【 0 1 1 7 】

ステップ 2 9 7 0 で処理を終了する。

【 0 1 1 8 】

図 1 7 に物理記憶位置管理・最適化部 4 2 が実施するデータ再配置処理の処理フローを示す。ここで、管理者の指示により処理を開始するモードと、あらかじめ設定されている時刻に自動的にデータ再配置案作成処理を実施し、その後

成されたデータ再配置案を実現するためにデータ移動を自動実行するデータ再配置自動実行モードの２種類を考える。

【 0 1 1 9 】

後述するように、複数の異なった種類のデータ配置解析・データ再配置案作成処理を実行可能であり、処理すべき種類の指定をして処理を開始する。また、処理にパラメータが必要な場合は併せてそれが指定されているものとする。これらは、管理者が処理を指示する場合にはそのときに一緒に指示を出し、データ再配置自動実行モードの場合には処理する種類や必要なパラメータを事前に設定しておく。

【 0 1 2 0 】

ステップ 2 0 0 1 でデータ再配置処理を開始する。このとき、データ配置解析・データ再配置案作成処理として何を実行するか指定する。また、必要であればパラメータを指定する。

【 0 1 2 1 】

ステップ 2 0 0 2 でデータ再配置処理に必要な DBMS データ情報 3 6 を前述した方法で取得し記憶する。なお、このデータ収集は、ステップ 2 0 0 1 の処理開始とは無関係にあらかじめ実行しておくこともできる。この場合には、情報を取得した時点から現在まで情報に変更がないかどうかをこのステップで確認する。

【 0 1 2 2 】

ステップ 2 0 0 3 では、ワーク領域を確保し、その初期化を行う。ワーク領域としては、図 1 8 に示すデータ再配置ワーク情報 6 7 0 と図 1 9 に示す移動プラン情報 7 5 0 を利用する。データ再配置ワーク情報 6 7 0 と移動プラン情報 7 5 0 の詳細とその初期データ作成方法は後述する。

【 0 1 2 3 】

ステップ 2 0 0 4 でデータ配置の解析・再配置案の作成処理を実行する。後述するように、データ配置の解析・再配置案作成処理は複数の観点による異なったものが存在し、このステップではステップ 2 0 0 1 で指定された処理を実行する。またステップ 2 0 0 1 でパラメータを受け取った場合には、それを実行する処

理に与える。

【0124】

ステップ2005ではステップ2004のデータ再配置案作成処理が成功したかどうか確認する。成功した場合にはステップ2006に進む。失敗した場合にはステップ2010に進み、管理者にデータ再配置案作成が失敗したことを通知し、ステップ2011に進み処理を完了する。

【0125】

ステップ2006では、現在データ再配置自動実行モードにより処理を実行しているか確認する。自動実行モードにより処理を実行している場合にはステップ2009に進む。そうでない場合には、ステップ2007に進む。

【0126】

ステップ2007では、ステップ2004で作成されたデータ再配置案を管理者に提示する。この提示を受けて管理者はデータ再配置案に問題がないか判断する。

ステップ2008では、データの再配置を続行するか否かを管理者から指示を受ける。続行する場合にはステップ2009に進み、そうでない場合にはステップ2011に進み処理を完了する。

【0127】

ステップ2009では、ステップ2004で作成されたデータの再配置案を基にデータの再配置処理を実行する。このとき、移動プラン情報750中の移動順序761で示される順に指定されたボリュームの領域を指定された物理記憶装置18内の領域へデータの移動を実行する。移動処理機能の実現方法は前述した通りである。

【0128】

ステップ2010でデータ再配置処理は完了である。

【0129】

図18はステップ2003において作成する情報であるデータ再配置ワーク情報670を示す。データ再配置ワーク情報670はデータ移動可能領域を保持する空き領域情報680とデータ構造物理記憶位置情報712のコピーを保持する

。空き領域情報 6 8 0 は、データ移動可能領域を示す物理記憶装置名 5 0 2 と物理ブロック番号 5 1 4 の組を保持する。

【 0 1 3 0 】

データの初期化は以下の方法で実行する。空き領域情報 6 8 0 はボリューム物理記憶位置メイン情報 5 1 0 中のボリューム名 5 0 1 が “E m p t y” である領域を集めることにより初期化する。データ構造物理記憶位置情報 7 1 2 は DB M S データ情報 3 6 に存在するデータをそのままコピーする。データ再配置案作成時にこれらのデータの値を変更するため、データ構造物理記憶位置情報 7 1 2 は必ずコピーを作成する。

【 0 1 3 1 】

図 1 9 はステップ 2 0 0 4 で実行されるデータ配置解析・データ再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報 7 5 0 を示す。移動プラン情報 7 5 0 は、移動指示の実行順序を示す移動順序 7 6 1、移動するデータを持つボリュームとそのデータ領域を示す移動ボリューム名 7 6 8 と移動ボリューム論理ブロック番号 7 6 9、そのデータの移動先の物理記憶装置とその記憶領域を示す移動先物理記憶装置名 7 7 1 と移動先物理ブロック番号 7 7 2 の組を保持する。この情報に関しては、何もデータを持たないように初期化する。

【 0 1 3 2 】

ステップ 2 0 0 4 で実行されるデータ配置解析・データ再配置案作成処理について説明する。前述のように、この処理には幾つかの種類が存在する。全ての処理に共通するのは逐次的にデータ再配置のためのデータ移動案を作成することである。そのため、データ移動の順番には意味があり、移動プラン情報 7 5 0 中の移動順序 7 6 1 にその順番を保持し、その順序どおりにデータ移動を行うことによりデータの再配置を実施する。

【 0 1 3 3 】

また、逐次処理のため、移動後のデータ配置をもとに次のデータの移動方法を決定する必要がある。そこで、データ移動案を作成するたびにデータ再配置ワーク情報 6 7 0 をデータ移動後の配置に更新する。

【 0 1 3 4 】

データ再配置案作成時のデータ移動案の作成は以下のように行う。移動したいデータ量以上の連続した移動可能領域をデータ再配置ワーク情報 6 7 0 中の情報から把握し、その中の領域を適当に選択し、設定条件や後述する制約を満たすかどうか確認をする。もし、それらを満たす場合にはそこを移動先として設定する。それらを満たさない場合には他の領域を選択し、再度それらを満たすかどうか確認をする。

【 0 1 3 5 】

以下、設定条件と制約を満たす領域を発見するか、全ての移動したいデータ量以上の連続した移動可能領域が設定条件や制約を満たさないことを確認するまで処理を繰り返す。もし、全ての領域で設定条件や制約を満たさない場合にはデータ移動案の作成に失敗として終了する。

【 0 1 3 6 】

このときに重要なのは移動後のデータ配置において、問題となる配置を行わないことである。特に R D B M S においては、特定のデータに関してはアクセスが同時に行われる可能性が高く、それらを異なる物理記憶装置 1 8 上に配置する必要がある。そこで、以下で説明する全てのデータの移動案を作成する場合には、移動するデータに含まれるデータ構造と、移動先に含まれるデータ構造を調べ、ログとその他のデータ、一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データがデータの移動後に同じ物理記憶装置 1 8 に配置されるかどうかを確認し、配置される場合には、その配置案は利用不可能と判断する。

【 0 1 3 7 】

なお、あるデータ構造がどの物理記憶装置 1 8 の領域に記憶されているか、また逆に、ある物理記憶装置 1 8 上の領域に記憶されるデータがどのデータ構造に対応するかは、データ再配置ワーク情報 6 7 0 中のデータ構造物理記憶位置情報 7 1 2 により把握可能である。

【 0 1 3 8 】

図 2 0 に第 1 のデータ配置解析・データ再配置案作成処理である、物理記憶装置稼動情報 3 2 を基にした同時アクセス実行データ構造を分離するためのデータ

再配置案作成処理の処理フローを示す。本処理においては、物理記憶装置 1 8 の稼働率が閾値を超えたものはディスクネック状態にあると判断してそれを解消するデータの移動案を作成する。本処理は、実測値に基づいて問題点を把握し、それを解決する方法を見つけるため、より精度の高いデータ再配置案を作成すると考えられ、データ再配置自動実行モードで最も利用しやすいものである。

【0 1 3 9】

ステップ 2 1 0 1 で処理を開始する。本処理を開始するにあたっては、どの期間の稼働率を参照するかを指定する。

【0 1 4 0】

ステップ 2 1 0 2 では、物理記憶装置 1 8 の識別子と指定期間における物理記憶装置 1 8 の稼働率の組を記憶するワーク領域を取得し、物理記憶装置稼働情報 3 2 を参照してその情報を設定し、物理記憶装置 1 8 の稼働率で降順にソートする。物理記憶装置稼働情報 3 2 中では、同じ物理記憶装置 1 8 中に記憶されているデータであっても異なるボリュームのものは分離して稼働率を取得しているため、それらの総和として物理記憶装置 1 8 の稼働率を求める必要がある。

【0 1 4 1】

ステップ 2 1 0 3 では、ステップ 2 1 0 2 のソート結果をもとに物理記憶装置 1 8 の稼働率が閾値を超えているもののリストである過負荷確認リストを作成する。このリスト中のエントリに関しても稼働率が降順になるような順序を保つようにする。

【0 1 4 2】

ステップ 2 1 0 4 では、過負荷確認リスト中にエントリが存在するか確認する。エントリが存在しない場合には、もう過負荷状態の物理記憶装置 1 8 が存在しないものとしてステップ 2 1 0 5 に進みデータ再配置案作成処理成功として処理を終了する。エントリが存在する場合には、ステップ 2 1 0 6 に進む。

【0 1 4 3】

ステップ 2 1 0 6 では、過負荷確認リスト中の最も物理記憶装置 1 8 の稼働率が高いものを再配置対象の物理記憶装置 1 8 として選択する。

【0 1 4 4】

ステップ 2 1 0 7 では、再配置対象となった物理記憶装置 1 8 内部のボリュームとその稼働率のリストを物理記憶装置稼働情報 3 2 を参照して作成し、稼働率で降順にソートする。

【 0 1 4 5 】

ステップ 2 1 0 8 では、リスト中のあるボリュームの稼働率があらかじめ定められた閾値を超過しているかどうか確認する。全てのボリュームの稼働率が閾値を超えていない場合には、ステップ 2 1 1 3 に進み、あるボリュームの稼働率がその閾値を超えている場合には、ステップ 2 1 0 9 に進む。

【 0 1 4 6 】

ステップ 2 1 0 9 においては、稼働率が閾値を超えているボリュームに関して、確認対象の物理記憶装置 1 8 中に同時にアクセスされる可能性があるデータの組、すなわち、ログとその他のデータ、一時表領域とその他のデータ、表データとそれに対して作成された木構造の索引データがあるそのボリューム内部に記憶されているかどうか発見する処理を行う。

【 0 1 4 7 】

ステップ 2 1 1 0 では、ステップ 2 1 0 9 における結果を確認し、同時アクセスデータ構造の組が存在する場合にはステップ 2 1 1 1 に進む。同時アクセスデータ構造の組が存在しない場合には、ステップ 2 1 1 2 に進む。

【 0 1 4 8 】

ステップ 2 1 1 1 においては、同時アクセスデータ構造の組に属するデータを異なる物理記憶装置 1 8 に記憶するためのデータ移動案を作成し、ステップ 2 1 1 4 に進む。

【 0 1 4 9 】

ステップ 2 1 1 2 においては、現在確認対象となっているボリューム内のデータを論理ブロック番号に従って 2 分割し、その片方を他の物理記憶装置 1 8 へ移動するデータ移動案を作成し、ステップ 2 1 1 4 に進む。

【 0 1 5 0 】

ステップ 2 1 1 3 においては、現在確認対象になっている物理記憶装置 1 8 の稼働率が閾値を下回るまで、稼働率が高いボリュームから順に、その物理記憶装

置 1 8 に記憶されているボリュームを構成するデータ全体を他の物理記憶装置 1 8 に移動するデータ移動案を作成し、ステップ 2 1 1 4 に進む。

【 0 1 5 1 】

ステップ 2 1 1 1, 2 1 1 2, 2 1 1 3 のデータ移動先を発見する際に、移動後の移動先の記憶装置の稼働率を予測する。物理記憶装置 1 8 毎の性能差が既知の場合にはその補正を行った移動データを含む記憶装置 1 8 上のボリュームの稼働率分、未知の場合には補正を行わない移動データを含む記憶装置 1 8 上のボリュームの稼働率分、データ移動により移動先の物理記憶装置 1 8 の稼働率が上昇すると考え、加算後の値が閾値を越えないような場所へのデータの移動案を作成する。

【 0 1 5 2 】

稼働率の加算分に関して、移動データ量の比率を考慮しても良いが、ここではデータ中のアクセスの偏りを考慮して移動データに全てのアクセスが集中したと考えた判断を行う。

【 0 1 5 3 】

ステップ 2 1 1 4 では、データ移動案の作成に成功したかどうかを確認し、失敗した場合にはステップ 2 1 1 7 に進みデータの再配置案作成処理失敗として処理を終了する。成功した場合にはステップ 2 1 1 5 に進む。

【 0 1 5 4 】

ステップ 2 1 1 5 では作成したデータ移動案を移動プラン情報 7 5 0 に追加し、ステップ 2 1 1 6 に進む。ステップ 2 1 1 6 ではデータ再配置ワーク情報 6 7 0 を作成したデータ移動案に従って修正し、移動先記憶装置 1 8 のステップ 2 1 0 2 で作成した物理記憶装置 1 8 毎の稼働情報の値を前述の移動後の稼働率判断値に修正する。その後、現在の確認対象の物理記憶装置 1 8 を過負荷確認リストから削除し、ステップ 2 1 0 4 に戻り次の確認を行う。

【 0 1 5 5 】

次に第 2 のデータ配置解析・データ再配置案作成処理である、クエリ実行同時アクセスデータ構造カウント情報 7 1 4 を用いた同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては

、クエリ実行同時アクセスデータ構造カウント情報 7 1 4 から同時にアクセスされるデータの組を取得し、それらを異なる物理記憶装置 1 8 に配置するデータ再配置案を作成する。

【0 1 5 6】

図 2 1 にクエリ実行時同時アクセスデータカウント情報 7 1 4 を利用する同時アクセス実行データ構造を分離するためのデータ再配置案作成処理処理フローを示す。ステップ 2 2 0 1 で処理を開始する。ステップ 2 2 0 2 において、カウント値 7 0 3 の全エントリの総和に対してカウント値 7 0 3 の値が一定割合以上のデータ構造とその所属する DBMS 1 1 0 a, 1 1 0 b の組を求め、それらを確認リストとして記憶する。

【0 1 5 7】

ステップ 2 2 0 3 でステップ 2 2 0 2 で求めた確認リスト中に含まれるデータ構造の組に関して、それらを異なる物理記憶装置 1 8 に記憶するデータ再配置案を作成し、ステップ 2 2 0 4 に進む。なお、ステップ 2 2 0 3 の処理に関しては、図 2 2 を用いて後で説明する。ステップ 2 2 0 4 では、ステップ 2 2 0 3 においてデータ再配置案の作成に成功したかどうかを確認し、成功した場合にはステップ 2 2 0 5 に進みデータ再配置案作成処理成功として処理を終了し、失敗した場合にはステップ 2 2 0 6 に進みデータ再配置案作成処理失敗として処理を終了する。

【0 1 5 8】

図 2 2 に指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す。本処理を開始するときには、データ構造名と物理記憶装置 1 8 から分離するデータ構造名の組のリストである確認リストを与える。

【0 1 5 9】

ステップ 2 3 0 1 で処理を開始する。ステップ 2 3 0 3 で確認リスト中にエントリが存在するか確認し、存在しない場合にはステップ 2 3 0 4 に進みデータ再配置案作成処理成功として処理を終了する。存在する場合にはステップ 2 3 0 5 に進む。

【0160】

ステップ2305においては、確認リストから1つ確認対象データ構造名とその所属DBMS名の組とその分離データ構造名とその所属DBMS名の組の組を取得し、ステップ2306に進む。

【0161】

ステップ2306においては、確認対象データ構造とその分離するデータ構造が同一の物理記憶装置上に記憶されているかどうかの確認を行う。この確認はデータ再配置ワーク情報670中のデータ構造物理記憶位置情報712を参照することにより可能である。両データ構造が全て異なる物理記憶装置上に存在する場合にはステップ2312に進み、ある物理記憶装置上に両データ構造が存在する場合にはステップ2307に進む。

【0162】

ステップ2307においては、同一の物理記憶装置上に両データ構造が存在する部分に関してそれを分離するデータ移動案を作成する。ステップ2308においては、そのデータ移動案作成が成功したかどうか確認し、成功した場合にはステップ2310に進み、失敗した場合にはステップ2309に進みデータ再配置案作成処理失敗として処理を終了する。

【0163】

ステップ2310においては、作成されたデータ移動案を移動プラン情報750に記憶する。ステップ2311においては、作成されたデータ移動案に従ってデータ再配置ワーク情報670を更新し、ステップ2312に進む。

【0164】

ステップ2312においては、確認リストから現在確認対象となっているデータ構造の組のエントリを削除し、2303に進む。

【0165】

図23に第3のデータ配置解析・データ再配置案作成処理である、データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す。本処理においては、同時にアクセスされる可能性が高い、ログとその他のデータ、一時表領域とその他のデータ、表データとそれ

に対して作成された木構造の索引データが同一物理記憶装置 1 8 上に記憶されている部分が存在しないか確認をし、そのような部分が存在する場合にはそれを解決するデータ再配置案を作成する。

【 0 1 6 6 】

ステップ 2 4 0 1 で処理を開始する。ステップ 2 4 0 2 では、DBMS データ構造情報 6 2 1 を参照して全てのログであるデータ構造名 5 6 1 とそれを利用する DBMS 1 1 0 a, 1 1 0 b の DBMS 名 6 3 1 の組を取得する。そして、そのデータ構造名とログ以外のデータを分離するデータ構造とする確認リストを作成し、ステップ 2 4 0 3 に進む。ステップ 2 4 0 3 ではステップ 2 4 0 2 で作成した確認リストを用いてステップ 2 3 0 1 から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。

【 0 1 6 7 】

ステップ 2 4 0 4 ではステップ 2 4 0 3 におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ 2 4 0 5 に進む。失敗した場合にはステップ 2 4 1 2 に進みデータ再配置案作成処理失敗として処理を終了する。

【 0 1 6 8 】

ステップ 2 4 0 5 では、DBMS データ構造情報 6 2 1 を参照して全ての一時表領域であるデータ構造名 5 6 1 とそれを利用する DBMS 1 1 0 a, 1 1 0 b の DBMS 名 6 3 1 の組を取得する。そして、そのデータ構造と一時表領域以外のデータを分離するデータ構造とする確認リストを作成し、ステップ 2 4 0 6 に進む。ステップ 2 4 0 6 ではステップ 2 4 0 5 で作成した確認リストを用いてステップ 2 3 0 1 から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。

【 0 1 6 9 】

ステップ 2 4 0 7 ではステップ 2 4 0 6 におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ 2 4 0 8 に進む。失敗した場合にはステップ 2 4 1 2 に進みデータ再配置案作成処理失敗として処理を終了する。

【 0 1 7 0 】

ステップ 2 4 0 8 では、DBMS 索引定義情報 6 2 4 を参照して全ての木構造

索引の索引名 6 3 5 とそれを利用する DBMS 1 1 0 a, 1 1 0 b の DBMS 名 6 3 1 の組とそれに対応する表のデータ構造名とそれを利用する DBMS 1 1 0 a, 1 1 0 b の DBMS 名 6 3 1 の組を対応表情報 6 3 7 から取得する。そして、それらの索引と表に関するデータを組とする確認リストを作成し、ステップ 2 4 0 9 に進む。ステップ 2 4 0 9 ではステップ 2 4 0 8 で作成した確認リストを用いてステップ 2 3 0 1 から開始されるデータ構造分離のためのデータ再配置案作成処理を実行する。ステップ 2 4 1 0 ではステップ 2 4 0 9 におけるデータ再配置案作成処理が成功したか確認をし、成功した場合にはステップ 2 4 1 1 に進み、データ再配置案作成処理成功として処理を終了する。失敗した場合にはステップ 2 4 1 2 に進みデータ再配置案作成処理失敗として処理を終了する。

【 0 1 7 1 】

図 2 4 に第 4 のデータ配置解析・データ再配置案作成処理である、特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す。この処理は、ランダムアクセス実行時の処理の並列度を考慮してディスクネックの軽減を図るためにデータの再配置を行うものである。この処理を実行する際には、データ再配置の確認対象とするデータ構造を DBMS 名 6 3 1 とデータ構造名 5 6 1 の組として指定する。

【 0 1 7 2 】

ステップ 2 5 0 1 で処理を開始する。ステップ 2 5 0 2 において、指定されたデータ構造の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、DBMS データ構造情報 6 2 1 中のデータ構造データ量 6 4 1 を参照することにより求める。

【 0 1 7 3 】

ステップ 2 5 0 3 においては、DBMS データ構造情報 6 2 1 を参照して指定データ構造における最大アクセス並列度 5 6 9 を取得する。

【 0 1 7 4 】

ステップ 2 5 0 4 において、ステップ 2 5 0 2 で求めた指定データ構造の記憶領域利用総量をステップ 2 5 0 3 で求めた最大アクセス並列度 5 6 9 で割った値を、指定データ構造の 1 つの物理記憶装置 1 8 上への割り当てを許可する最大量

として求める。この制約により、特定の物理記憶装置 1 8 に偏ることなく最大アクセス並列度 5 6 9 以上の台数の物理記憶装置 1 8 に指定データ構造が分散して記憶されることになり、最大アクセス並列度 5 6 9 による並列度でランダムアクセスが実行されてもディスクネックになりにくい状況となる。なお、割り当て許可最大量の値は、実際のアクセス特性を考慮してこの方法で求めた値から更に増減させても構わない。

【 0 1 7 5 】

ステップ 2 5 0 5 において、指定データ構造のデータがステップ 2 5 0 4 で求めた最大量を超えて 1 つの物理記憶装置 1 8 上に割り当てられているものが存在するかデータ再配置ワーク情報 6 7 0 を用いて確認し、そのようなものが存在しない場合にはステップ 2 5 0 9 に進み、データ再配置案作成処理成功として処理を終了する。存在する場合にはステップ 2 5 0 6 に進む。

【 0 1 7 6 】

ステップ 2 5 0 6 においては、ステップ 2 5 0 4 で求めた最大量を超えて 1 つの物理記憶装置 1 8 上に割り当てられている部分を解消するデータ移動案を作成する。このとき、移動案作成に考慮するデータ移動量は指定データ構造の現在の物理記憶装置 1 8 上への割り当て量のステップ 2 5 0 4 で求めた最大量からの超過分以上である必要がある。また、移動先物理記憶装置 1 8 においても、移動後にステップ 2 5 0 4 で求めた最大量を超過しないようにする必要がある。

【 0 1 7 7 】

ステップ 2 5 0 7 においては、ステップ 2 5 0 6 のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ 2 5 0 8 に進む。失敗した場合にはステップ 2 5 1 0 に進み、データ再配置案作成処理失敗として処理を終了する。

【 0 1 7 8 】

ステップ 2 5 0 8 においては作成したデータ移動案を移動プラン情報 7 5 0 に記憶し、ステップ 2 5 0 9 に進みデータ再配置案作成処理成功として処理を終了する。

【 0 1 7 9 】

図 2 5 に第 5 のデータ配置解析・データ再配置案作成処理である、特定の表データに対するシーケンシャルアクセス時のディスクネックを解消するデータ再配置案作成処理の処理フローを示す。この処理を実行する際には、データ再配置の確認対象とする表を DBMS 名 6 3 1 とデータ構造名 5 6 1 の組として指定する。

【 0 1 8 0 】

前述のように、対象とする DBMS 1 1 0 a, 1 1 0 b の種類が絞られるが、データ構造物理記憶位置情報 7 1 2 はシーケンシャルアクセス順にソートされてデータを記憶しているため、シーケンシャルアクセス方法は既知である。

【 0 1 8 1 】

また、並列にシーケンシャルアクセスを実行する場合に、その領域の分割法は並列にアクセスしない場合のシーケンシャルにアクセスする順番を並列度に合わせて等分に分割するものとする。

【 0 1 8 2 】

この並列アクセスによる分割後の 1 つのアクセス領域を全て同一の物理記憶装置 1 8 上に配置するのは必ずしも現実的ではない。そこで、分割後のアクセス領域がある一定量以上連続にまとまって 1 つの物理記憶装置上に記憶されていればよいと判断する。ただし、どのような場合でも連続してアクセスされることがなく、分割後のアクセス領域が異なるものに分類されるものに関しては、並列シーケンシャルアクセス時にアクセスがぶつかる可能性があるため、異なる物理記憶装置 1 8 に記憶するという指針を設けて、これに沿うようなデータ配置を作成することによりシーケンシャルアクセスの性能を高める。

【 0 1 8 3 】

ステップ 2 6 0 1 で処理を開始する。ステップ 2 6 0 2 において、指定された表の物理記憶装置上に割り当てられた記憶領域利用総量を求める。この値は、DBMS データ構造情報 6 2 1 中のデータ構造データ量 6 4 1 を参照することにより求める。ステップ 2 6 0 3 においては、DBMS データ構造情報 6 2 1 を参照して指定データ構造における最大アクセス並列度 5 6 9 を取得する。

【 0 1 8 4 】

ステップ 2 6 0 4 において、ステップ 2 6 0 2 で求めた指定表の記憶領域利用総量をステップ 2 6 0 3 で求めた最大アクセス並列度 5 6 9 で割った量が、並列アクセス時にシーケンシャルにアクセスされる 1 つの領域のデータ量である。データ構造物理記憶位置情報 7 1 2 はシーケンシャルアクセス実行順にソートされているため、これを用いて最大アクセス並列度 5 6 9 の並列アクセスが実行されると仮定した前述のデータ分割指針を作成する。

【 0 1 8 5 】

ステップ 2 6 0 5 において、データ再配置ワーク情報 6 7 0 を参照しながら、指定データ構造はステップ 2 6 0 4 で作成したデータ分割指針に沿ったデータ配置が物理記憶装置 1 8 上で行われているか確認し、そうであればステップ 2 6 0 9 に進み、データ再配置案作成処理成功として処理を終了する。そうでない場合にはステップ 2 6 0 6 に進む。

【 0 1 8 6 】

ステップ 2 6 0 6 においては、物理記憶装置 1 8 上において、ステップ 2 6 0 4 で求めたデータ分割指針に従うデータ配置を求める。このとき、データがある一定値以下の領域に細分化されている場合には、連続した空き領域を探し、そこにアクセス構造を保つようにデータを移動するデータ移動案を作成する。また、最大アクセス並列度 5 6 9 の並列アクセスにより異なるアクセス領域に分離されるデータが同じ物理記憶装置 1 8 上に配置されないようなデータ移動案を作成する。

【 0 1 8 7 】

ステップ 2 6 0 7 においては、ステップ 2 6 0 6 のデータ移動案作成処理が成功したか確認をする。成功した場合にはステップ 2 6 0 8 に進み、失敗した場合にはステップ 2 6 1 0 に進み、データ再配置案作成処理失敗として処理を終了する。

【 0 1 8 8 】

ステップ 2 6 0 8 においては作成したデータ移動案を移動プラン情報 7 5 0 に記憶し、ステップ 2 6 0 9 に進みデータ再配置案作成処理成功として処理を終了する。

＜第二の実施の形態＞

本実施の形態では、DBMSが実行される計算機とファイルを管理単位とする記憶装置がネットワークを用いて接続された計算機システムにおいて、記憶装置がDBMSに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報を取得し、それらを用いて記憶装置の動作を改善する。

【0189】

記憶装置において、記憶装置内部でデータの記憶位置を動的に変更する機能を有し、取得した情報をもとに好適なデータ再配置案を作成し、データの記憶位置の動的変更機能を用いて、作成したデータ再配置案に従ったデータ配置を実現し、アクセス性能を改善する。また、取得情報をもとにしたデータキャッシュの制御を行いより良いアクセス性能特性が得られるようにする。

【0190】

図26は、本発明の第二の実施の形態における計算機システムの構成図である。図示されたように、本発明の第二の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【0191】

本実施の形態においてはI/Oパスインターフェイス70、I/Oパス71、I/Oパススイッチ72が存在せず、記憶制御装置10bとDBホスト80c、80dはネットワーク79を介してのみ接続される。記憶装置10はファイルを単位としたデータ記憶管理を行う記憶装置10bに変更される。そのため、物理記憶装置稼動情報32、データキャッシュ管理情報34、DBMSデータ情報36、ボリューム物理記憶位置管理情報38がそれぞれ物理記憶装置稼動情報32b、データキャッシュ管理情報34b、DBMSデータ情報36b、ファイル記憶管理情報38bに変更される。

【0192】

DBホスト80c、80dで実行されるOS100ではボリュームマネージャ102、ファイルシステム104が削除されその代わりに記憶装置10bが提供するファイルをアクセスするための機能を有するネットワークファイルシステム104bが追加され、OS100が保持するマッピング情報106がネットワー

クマウント情報106bへ変更される。

【0193】

記憶装置10はファイルを管理単位とする記憶装置10bに変更される。DBホスト80c, 80dからのアクセスもNFS等のファイルをベースとしたプロトコルで実施される。記憶装置10におけるボリュームの役割は、記憶装置10bにおいてはファイルもしくはファイルを管理するファイルシステムとなり、そのファイルの記憶位置管理情報がファイル記憶管理情報38bである。1つの記憶装置10bの中に複数のファイルシステムが存在しても構わない。物理記憶装置18の稼動情報はボリュームを単位とした取得からファイルシステムまたはファイルを単位とした取得に変更する。記憶装置10b内にファイルシステムが存在する場合でもデータの移動機能を実現可能である。

【0194】

図27はDBホスト80c, 80dのOS100内に記憶されているネットワークマウント情報106bを示す。ネットワークマウント情報106bは、記憶装置10bから提供され、DBホスト80c, 80dにおいてマウントされているファイルシステムの情報で、ファイルシステムの提供元記憶装置とそのファイルシステムの識別子である記憶装置名583とファイルシステム名1001、そして、そのファイルシステムのマウントポイントの情報であるマウントポイント1031の組を保持する。

【0195】

図28は記憶装置10b内に保持されるファイル記憶管理情報38bを示す。図5のボリューム物理記憶位置管理情報38からの変更点は、ボリューム物理記憶位置メイン情報510、ボリュームデータ移動管理情報511からファイル物理記憶位置情報510b、ファイルデータ移動管理情報511bにそれぞれ変更される。上記の変更内容は、ボリュームの識別子であるボリューム名501がファイルの識別子となるファイルシステム名1001とファイルパス名1002に、ボリューム内のデータ領域を示すボリューム論理ブロック番号512と移動論理ブロック番号782がそれぞれファイルブロック番号1003または移動ファイルブロック番号1021に変更されるものである。

【 0 1 9 6 】

ここで、ファイルパス名 1 0 0 2 が “ E m p t y ” であるエントリ 1 0 1 5 は特殊なエントリであり、このエントリには記憶装置 1 0 b 内の物理記憶装置 1 8 の領域のうち、指定ファイルシステム内でファイルの記憶領域として割り当てられていない領域を示し、図 5 中のボリュームデータ移動管理情報 5 1 1 を用いるデータ移動方式で説明した処理手順を用い、この領域に対して移動するデータをコピーすることによりデータの物理記憶位置の動的変更機能を実現する。

【 0 1 9 7 】

ここで注意が必要なのは、データ移動案作成時にデータ移動先の制約が増えた点である。本実施の形態においては、ファイルシステムを複数保持することが許されている。一般のファイルシステムにおいては、あるファイルシステムが他のファイルシステムが管理する領域を利用することは不可能である。つまり、一般のファイルシステムを用いている場合には、ファイルの移動は、そのファイルが存在しているファイルシステム内に閉じる必要がある。ただし、あるファイルシステムが他のファイルシステムが管理する領域を利用可能な機構を有している場合にはこの限りではない。

【 0 1 9 8 】

図 2 9 に記憶装置 1 0 b 内に保持される物理記憶装置稼動情報 3 2 b を示す。図 6 の物理記憶装置稼動情報 3 2 からの変更点は、稼動情報取得単位がボリュームからファイルシステムに変更されたため、ボリューム名 5 0 1 の部分がファイルシステム名 1 0 0 1 に変更されたことである。また、稼動情報取得単位をファイルとしてもよく、このときはボリューム名 5 0 1 の部分がファイルシステム名 1 0 0 1 とファイルパス名 1 0 0 2 に変更される。

【 0 1 9 9 】

図 3 0 に記憶装置 1 0 b 内に保持されている DBMS データ情報 3 6 b を示す。図 7 の DBMS データ情報 3 6 からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたためデータ構造物理記憶位置情報 7 1 2 に修正が加えられ、データ構造物理記憶位置情報 7 1 2 b に変更されたことである。

【0200】

図31にDBMSデータ情報36b中に含まれるデータ構造物理記憶位置情報712bを示す。図9のデータ構造物理記憶位置情報712からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、ボリューム名501とボリューム論理ブロック番号512の部分がファイルシステム名1001とファイルパス名1002とファイルブロック番号1003に変更されたことである。この情報は、DBMSデータ記憶位置情報622とネットワークマウント情報106bを記憶装置10の外部から取得し、さらにファイル物理記憶位置情報510bを参照して、対応する部分を組み合わせることにより作成する。

【0201】

図32に記憶装置10b内に保持されているデータキャッシュ管理情報34bを示す。図13のデータキャッシュ管理情報34からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、キャッシュセグメント情報720に修正が加えられ、キャッシュセグメント情報720bに変更されたことである。キャッシュセグメント情報720bのキャッシュセグメント情報720からの変更点は、上述の理由により、ボリューム名501とボリューム論理ブロック番号512の部分がファイルシステム名1001とファイルパス名1002とファイルブロック番号1003に変更されたことである。

【0202】

図33にステップ2003において作成する情報であるデータ再配置ワーク情報670bを示す。図18のデータ再配置ワーク情報670からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、空き領域情報680とデータ構造物理記憶位置情報712に修正が加えられ、それぞれ空き領域情報680bとデータ構造物理記憶位置情報712bへ変更されたことである。空き領域情報680bの空き領域情報680からの変更点は、ファイルシステムを利用した領域管理を実施しているため、空き領域管理はファイルシステムを意識する必要があるため、空き領域情報としては、データの記憶に利用していない場所を示す物理記憶装置名502と物理ブロック番号514とその空

き領域を管理するファイルシステム名 1 0 0 1 の組を保持する。空き領域情報 6 8 0 b はファイル物理記憶位置メイン情報 5 1 0 b 中のファイルパス名 1 0 0 1 が “E m p t y” である領域を集めることにより初期化する。

【 0 2 0 3 】

図 3 4 はステップ 2 0 0 4 で実行されるデータ配置解析・データ再配置案作成処理により作成されるデータ移動案を格納する移動プラン情報 7 5 0 b を示す。

図 1 9 の移動プラン情報 7 5 0 からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、移動ボリューム名 5 6 8 と移動ボリューム論理ブロック番号 7 6 9 の部分が移動ファイルシステム名 1 1 0 1 と移動ファイルパス名 1 1 0 2 と移動ファイルブロック番号 1 1 0 3 に変更されたことである。

【 0 2 0 4 】

前述のように、記憶装置 1 0 において一般のファイルシステムを用いている場合には、ファイルの移動は、そのファイルが存在しているファイルシステム内に閉じる必要がある。従って、データ再配置案作成処理に関して、本実施の形態における第一の実施の形態から変更点は、データ移動先は現在データが存在しているファイルシステム上に限られるという制約が追加される。ただし、この制約も記憶装置 1 0 が利用しているファイルシステムが他のファイルシステムが管理する領域を利用可能な機構を有している場合には除かれる。

【 0 2 0 5 】

記憶装置 1 0 b における本実施の形態における第一の実施の形態からの差は、ほとんどがボリューム名 5 0 1 をファイルシステム名 1 0 0 1 とファイルパス名 1 0 0 2 に、ボリューム論理ブロック番号 5 1 2 をファイルブロック番号 1 0 0 3 に変更することであり、その他の変更点もその差を述べてきた。記憶装置 1 0 b における処理に関しては、前記のデータ再配置案作成処理における制約を除き、基本的にこれまで述べてきた変更点と同じ変更点への対応方法を実施すれば、第一の実施の形態における処理をほぼそのまま本実施の形態に当てはめることができる。

【 0 2 0 6 】

【発明の効果】

本発明により以下のことが可能となる。第一に、DBMSが管理するデータを保持する記憶装置において、DBMSの処理の特徴を考慮することによりDBMSに対してより好ましい性能特性を持つことができる。この記憶装置を用いることにより、既存のDBMSに対してプログラムの修正無しにDBMS稼動システムの性能を向上させることができるようになる。つまり、高性能なDBシステムを容易に構築できるようになる。

【0207】

第二に、記憶装置の性能最適化機能を提供するため、それにより記憶装置の性能に関する管理コストを削減することができる。特に、本発明は、DBシステムの高性能化に寄与するため、この記憶装置を用いたDBシステムの性能に関する管理コストを削減することができる。更に、本発明を用いた記憶装置は、自動でDBMSの特性を考慮したデータ配置の改善を行うことができ、管理コストの削減に大きく寄与する。

【図面の簡単な説明】

【図1】

第一の実施の形態における計算機システムの構成を示す図である。

【図2】

DBホスト80a, 80bのOS100内に記憶されているマッピング情報106を示す図である。

【図3】

DBMS110a, 110b内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報114を示す図である。

【図4】

DBホスト80a, 80bのメモリ88上に記憶されている実行履歴情報122を示す図である。

【図5】

記憶装置10内に保持されているボリューム物理記憶位置管理情報38を示す図である。

【図 6】

記憶装置 1 0 内に保持されている物理記憶装置稼動情報 3 2 を示す図である。

【図 7】

記憶装置 1 0 内に保持されている DBMS データ情報 3 6 を示す図である。

【図 8】

DBMS データ情報 3 6 中に含まれる DBMS スキーマ情報 7 1 1 を示す図である。

【図 9】

DBMS データ情報 3 6 中に含まれるデータ構造物理記憶位置情報 7 1 2 を示す図である。

【図 1 0】

DBMS データ情報 3 6 中に含まれるクエリ実行同時アクセスデータ構造カウンタ情報 7 1 4 を示す図である。

【図 1 1】

DBMS データ情報 3 6 に含まれる DBMS データ構造キャッシュ効果情報 7 1 5 を示す図である。

【図 1 2】

記憶装置 1 0 において指定されたデータ構造のデータをデータキャッシュに保持する効果があるかどうかの判断する処理のフローを示す図である。

【図 1 3】

記憶装置 1 0 内に保持されているデータキャッシュ管理情報 3 4 を示す図である。

【図 1 4】

記憶装置 1 0 がホストからデータの読出し要求を受け取ったときの処理フローを示す図である。

【図 1 5】

記憶装置 1 0 がホストからデータの書き込み要求を受け取ったときの処理フローを示す図である。

【図 1 6】

アクセス先のデータの内容に従いアクセス要求のあったデータを保持するセグメントを適当な管理リストに繋ぐ処理のフローを示す図である。

【図 1 7】

記憶装置 1 0 内で実施されるデータ再配置処理の処理フローを示す図である。

【図 1 8】

データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報 6 7 0 を示す図である。

【図 1 9】

データ配置解析・再配置案作成処理で作成されるデータ移動案を格納する移動プラン情報 7 5 0 を示す図である。

【図 2 0】

物理記憶装置稼動情報 3 2 を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。

【図 2 1】

クエリ実行時同時アクセスデータカウント情報 7 1 4 を利用する同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。

【図 2 2】

指定されたデータ構造とそのデータ構造と同時にアクセスされる可能性が高いデータ構造の組を分離するデータ再配置案を作成する処理のフローを示す図である。

【図 2 3】

データ構造の定義を基にした同時アクセス実行データ構造を分離するためのデータ再配置案作成処理の処理フローを示す図である。

【図 2 4】

特定の表や索引の同一データ構造に対するアクセス並列度を考慮したデータ再配置案作成処理の処理フローを示す図である。

【図 2 5】

特定の表データに対するシーケンシャルアクセス時のディスクネックを解消す

るデータ再配置案作成処理の処理フローを示す図である。

【図 2 6】

第二の実施の形態における計算機システムの構成を示す図である。

【図 2 7】

DBホスト 8 0 c, 8 0 d の OS 1 0 0 内に記憶されているネットワークマウント情報 1 0 6 b を示す図である。

【図 2 8】

記憶装置 1 0 b 内に保持されるファイル記憶管理情報 3 8 b を示す図である。

【図 2 9】

記憶装置 1 0 b 内に保持される物理記憶装置稼動情報 3 2 b を示す図である。

【図 3 0】

記憶装置 1 0 b 内に保持されている DBMS データ情報 3 6 b を示す図である。

【図 3 1】

DBMS データ情報 3 6 b 中に含まれるデータ構造物理記憶位置情報 7 1 2 b を示す図である。

【図 3 2】

記憶装置 1 0 b 内に保持されているデータキャッシュ管理情報 3 4 b を示す図である。

【図 3 3】

データ配置解析・再配置案作成処理で利用するデータ再配置ワーク情報 6 7 0 b を示す図である。

【図 3 4】

データ配置解析・再配置案作成処理で作成されるデータ移動案を格納する移動プラン情報 7 5 0 b を示す図である。

【符号の説明】

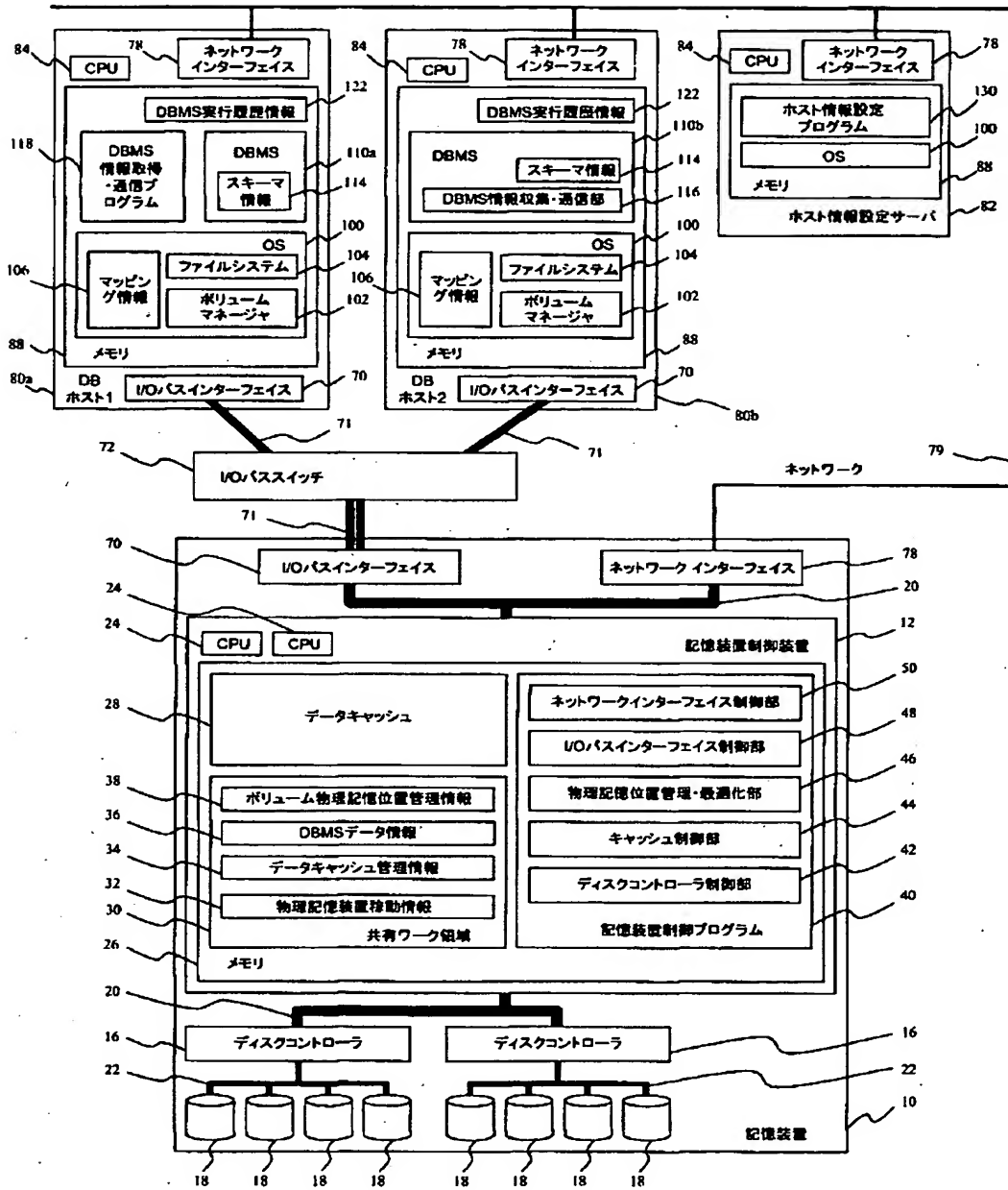
1 0, 1 0 b	記憶装置
1 8	物理記憶装置
2 8	データキャッシュ

3 2 , 3 2 b	物理記憶装置稼動情報
3 4 , 3 4 b	データキャッシュ管理情報
3 6 , 3 6 b	DBMSデータ情報
3 8	ボリューム物理記憶位置管理情報
3 8 b	ファイル記憶管理情報
4 0	記憶装置制御プログラム
4 2	ディスクコントローラ制御部
4 4	キャッシュ制御部
4 6	物理記憶位置管理・最適化部
4 8	I/Oバスインターフェイス制御部
5 0	ネットワークインターフェイス制御部
7 0	I/Oバスインターフェイス
7 1	I/Oバス
7 2	I/Oバススイッチ
7 8	ネットワークインターフェイス
7 9	ネットワーク
8 0 a , 8 0 b , 8 0 c , 8 0 d	DBホスト
8 2	ホスト情報設定サーバ
1 0 0	OS (オペレーティングシステム)
1 0 2	ボリュームマネージャ
1 0 4	ファイルシステム
1 0 4 b	ネットワークファイルシステム
1 0 6	マッピング情報
1 0 6 b	ネットワークマウント情報
1 1 0 a , 1 1 0 b	DBMS (データベース管理システム)
1 1 4	スキーマ情報
1 1 6	DBMS 情報通信部
1 1 8	DBMS 情報取得・通信プログラム
1 2 2	実行履歴情報

【書類名】 図面

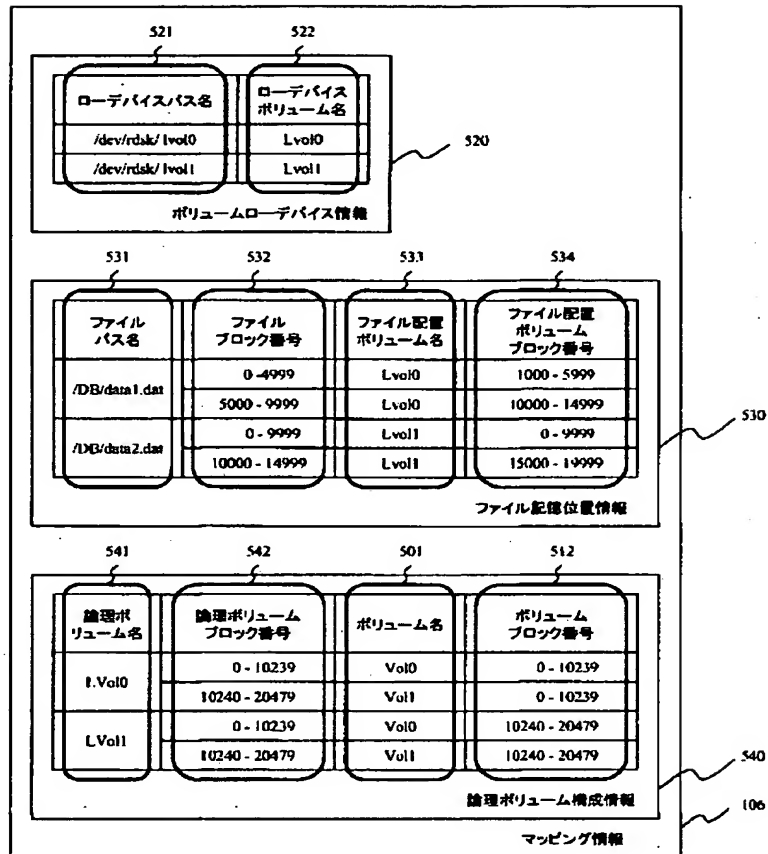
【図 1】

図 1



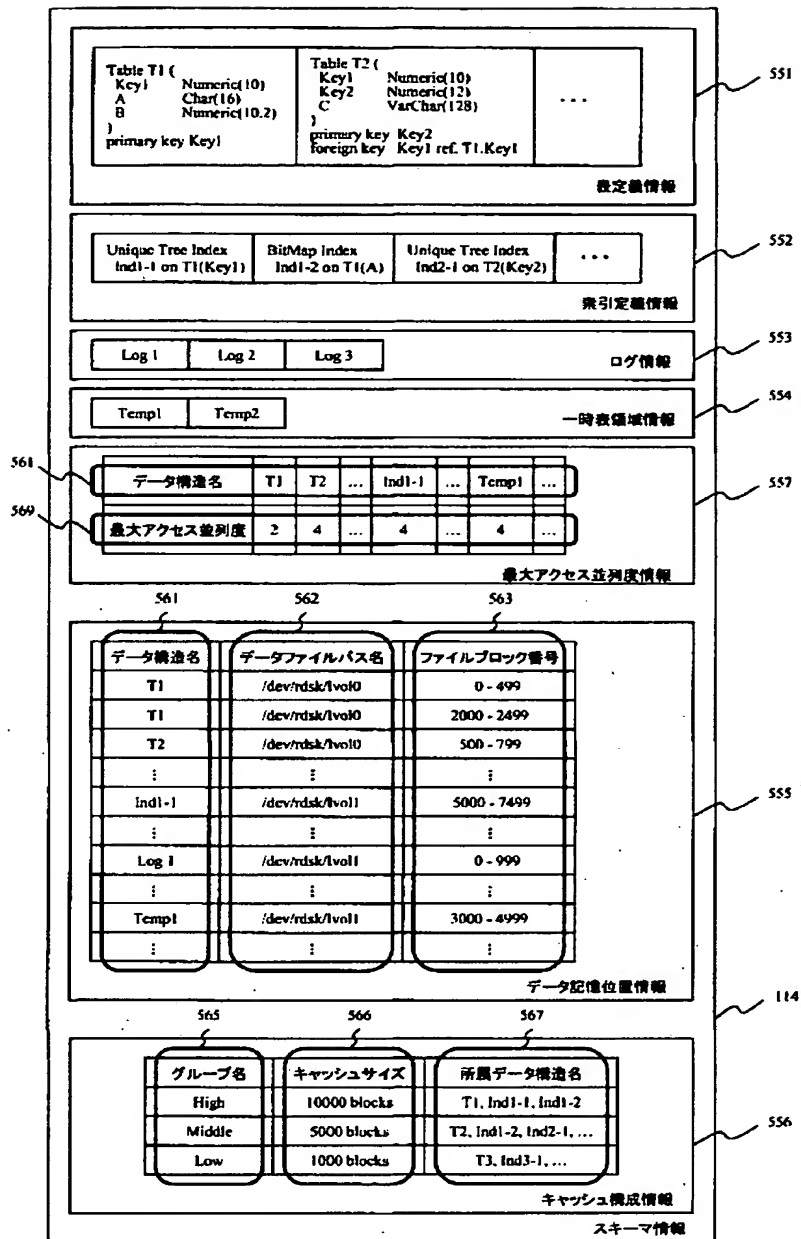
【図 2】

図2



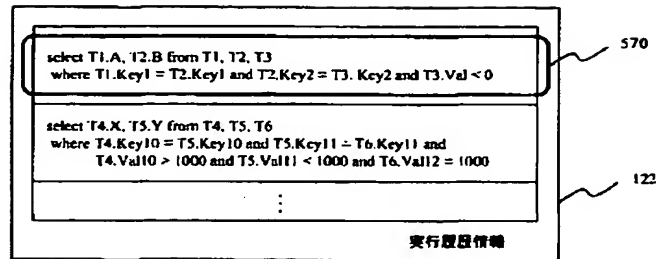
【図 3】

図3



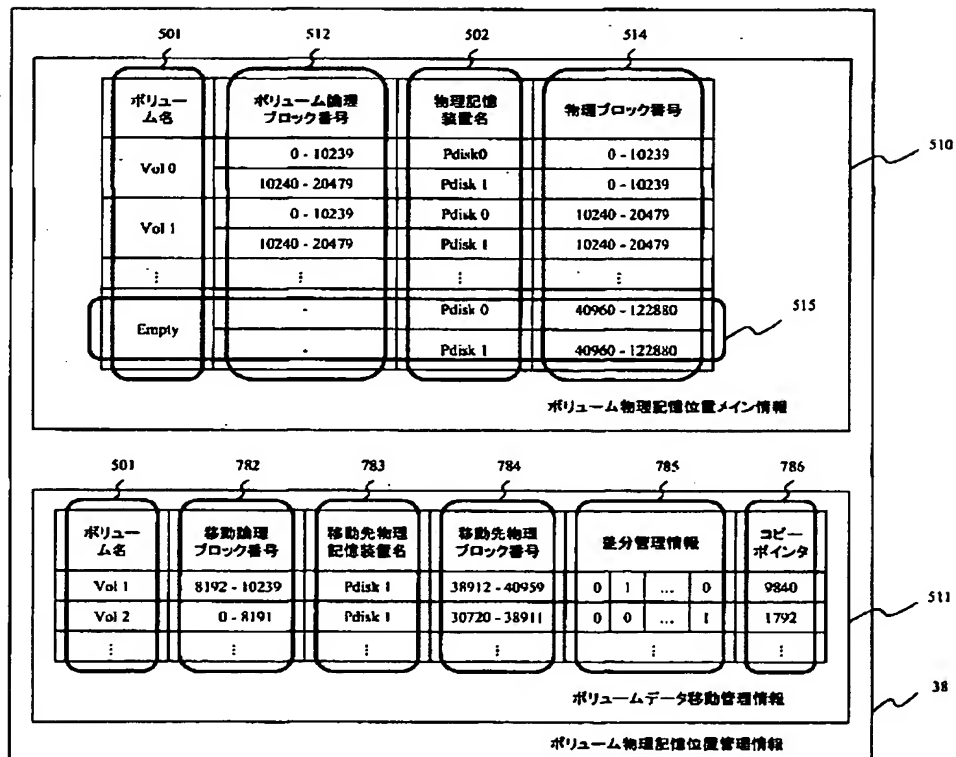
【図 4】

図4



【図 5】

図5



【図 6】

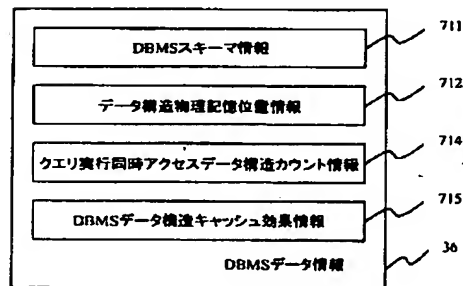
図6

ボリューム名		Vol0	Vol0	Vol1	...	501
物理記憶装置名		Pdisk 0	Pdisk 1	Pdisk 0	...	502
累積移動時間		23917390	38902849	8012891	...	503
旧累積移動時間		22787638	38783484	7592039	...	593
率 備 注	2000/4/1 12:00 ~ 2000/4/1 12:15	20%	12%	4%	...	594
	2000/4/1 12:15 ~ 2000/4/1 12:30	15%	10%	7%	...	
	2000/4/1 12:30 ~ 2000/4/1 12:45	16%	9%	5%	...	
	⋮	⋮	⋮	⋮	⋮	
	⋮	⋮	⋮	⋮	⋮	
前回累積移動時間取得時刻: 2001/4/12 18:15		物理記憶装置移動情報				32

595

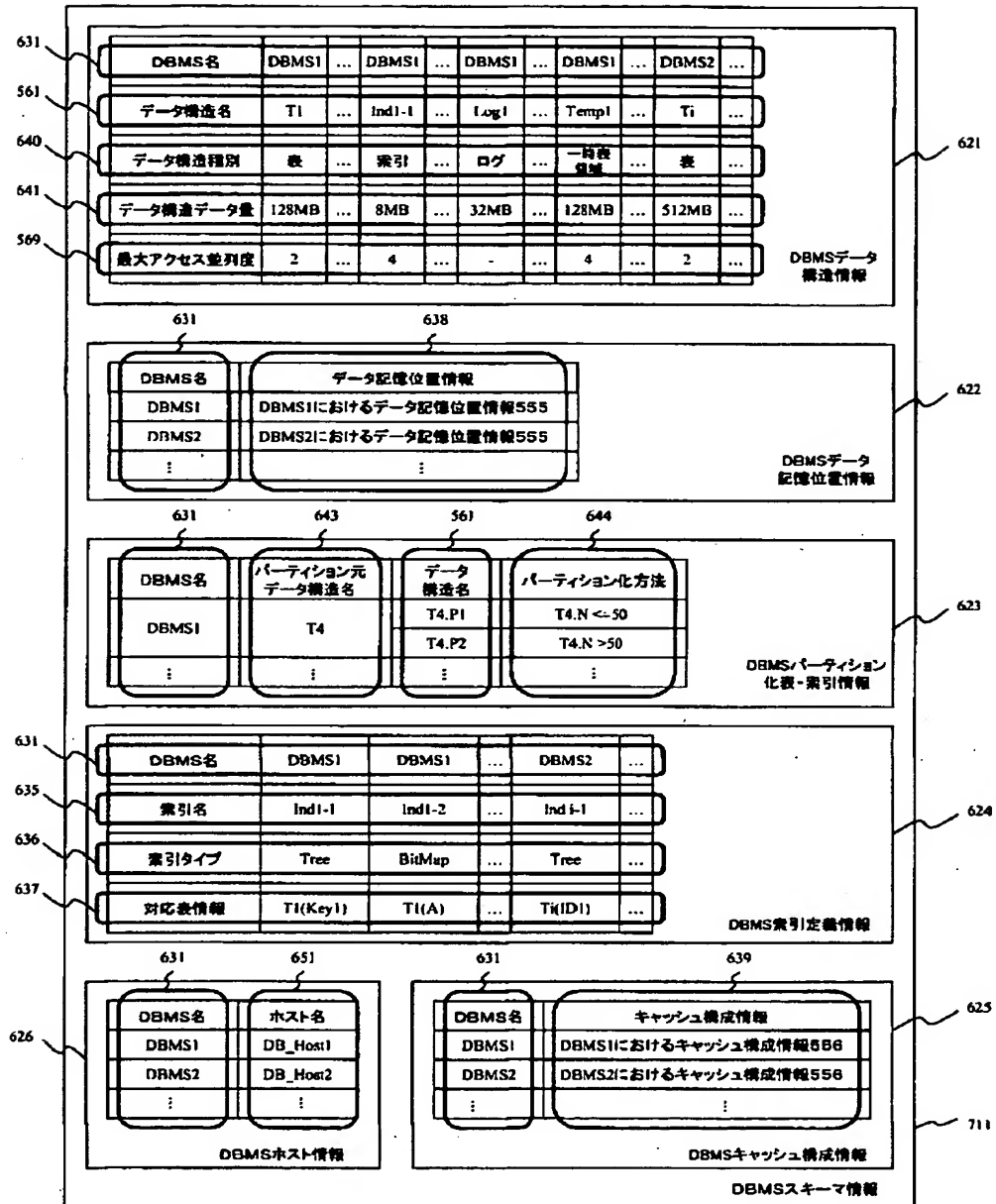
【図 7】

図7



【図8】

図8



【図 9】

図 9

DBMS名	DBMS1	DBMS1	DBMS1	...	DBMS2	...	631
データ構造名	T1	T1	T2	...	Ti	...	561
ボリューム名	Vol0	Vol1	Vol0	...	Vol2	...	501
ボリュームブロック番号	0 - 4999	5000 - 9999	0 - 239	...	0 - 9999	...	512
物理記憶装置名	Pdisk 0	Pdisk 0	Pdisk 0	...	Pdisk 1	...	502
物理ブロック番号	0 - 4999	10240 - 15239	10000 - 10239	...	20480 - 30479	...	514
データ構造物理記憶位置情報							712

【図 1 0】

図 10

631	701	702	703	
DBMS名	データ構造名A	データ構造名B	カウント値	
DBMS1	T1	Ind1-1	2789	
⋮	⋮	⋮	⋮	
DBMS2	Ti	Ind i-1	829	714
⋮	⋮	⋮	⋮	
クエリ実行同時アクセスデータ構造カウント情報				

【図 1 1】

図 1 1

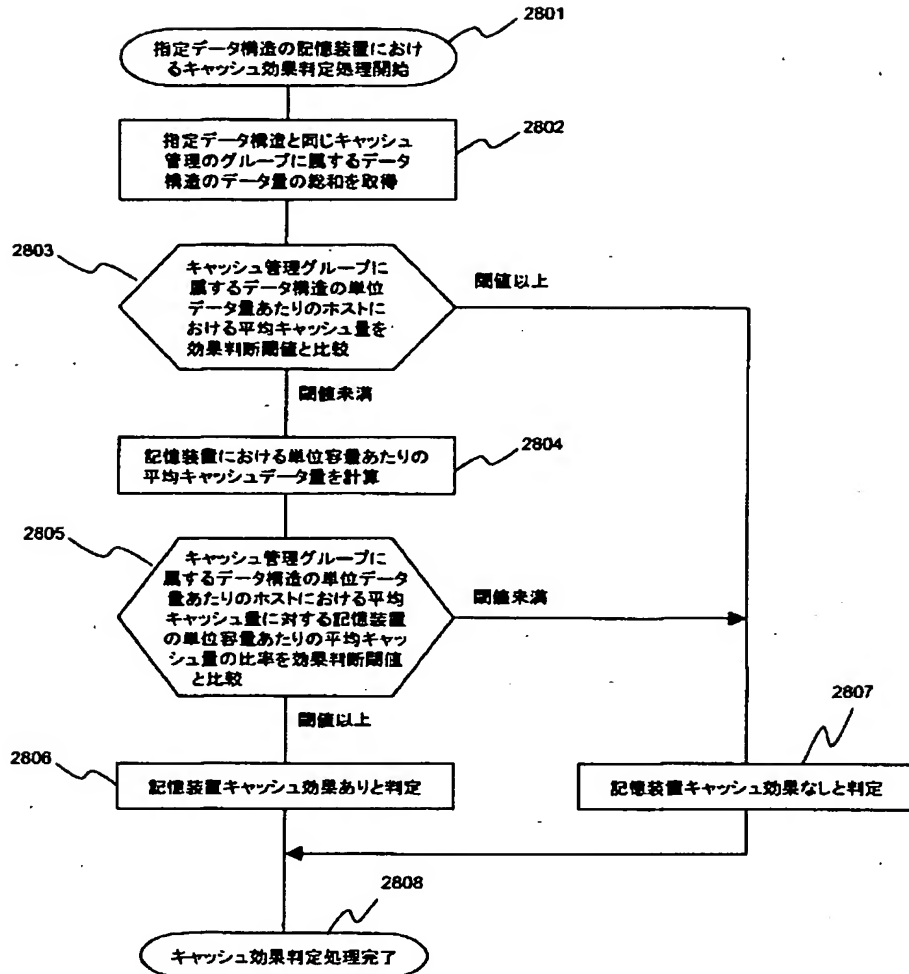
631 DBMS名	561 データ構造名	733 キャッシュ 効果情報
DBMS1	T1	あり
DBMS1	Ind1-1	あり
DBMS1	T2	なし
⋮	⋮	⋮
DBMS2	Ti	あり
⋮	⋮	⋮

DBMSデータ構造キャッシュ効果情報

715

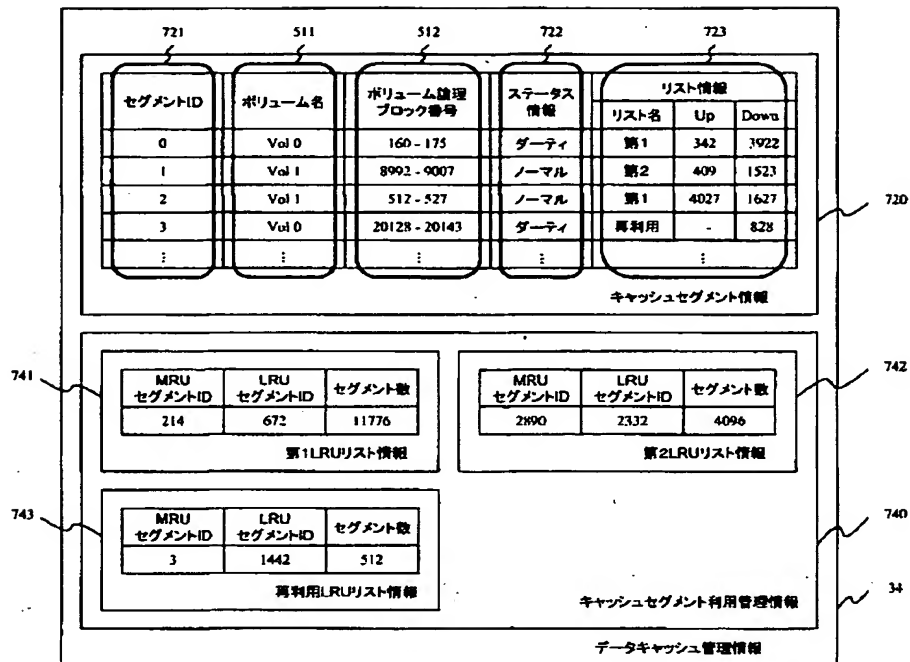
【図 12】

図 12



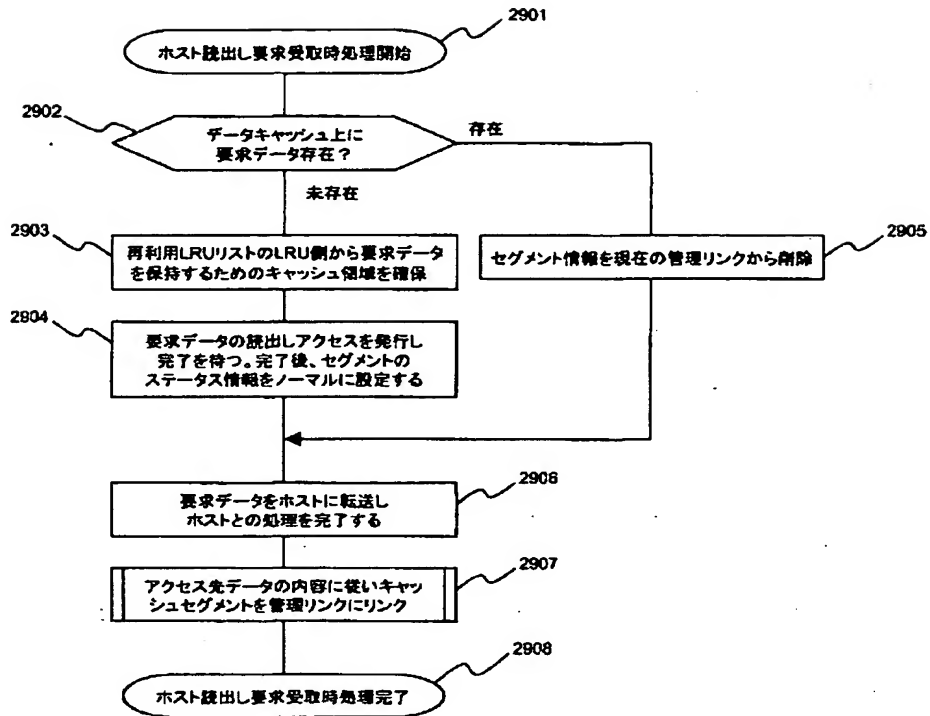
【図 13】

図13



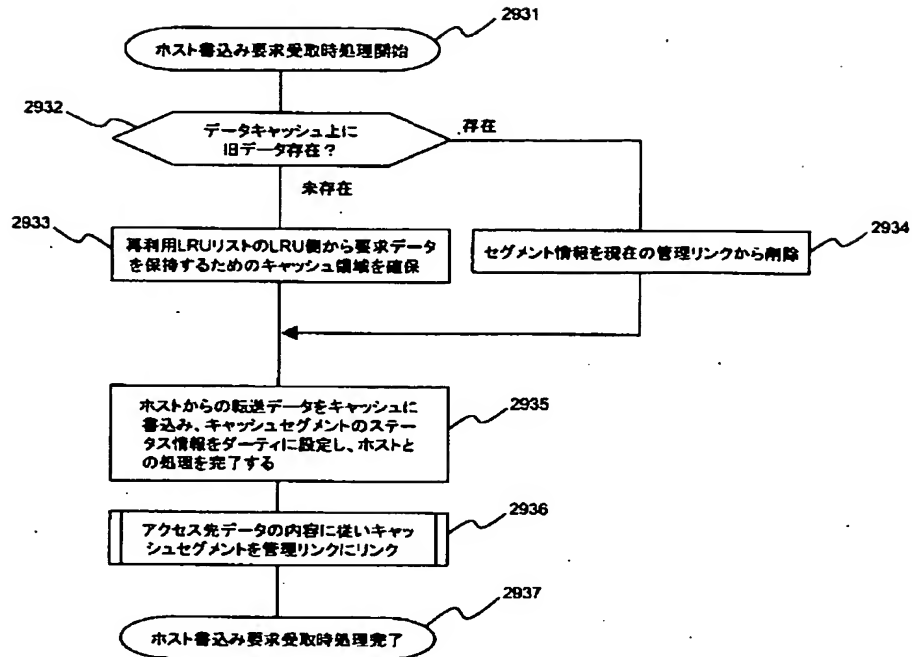
【図 14】

図 14

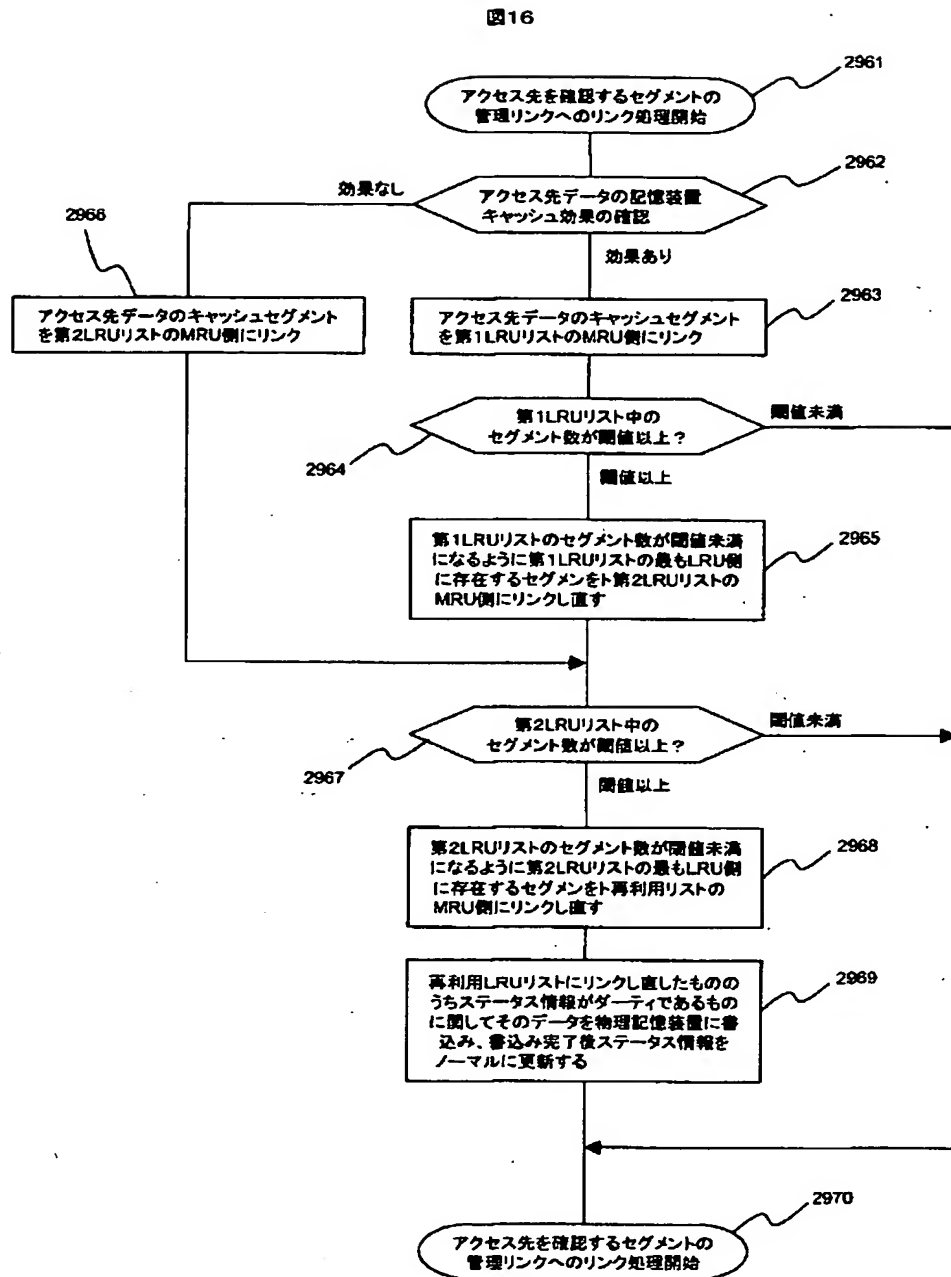


【図 15】

図 16

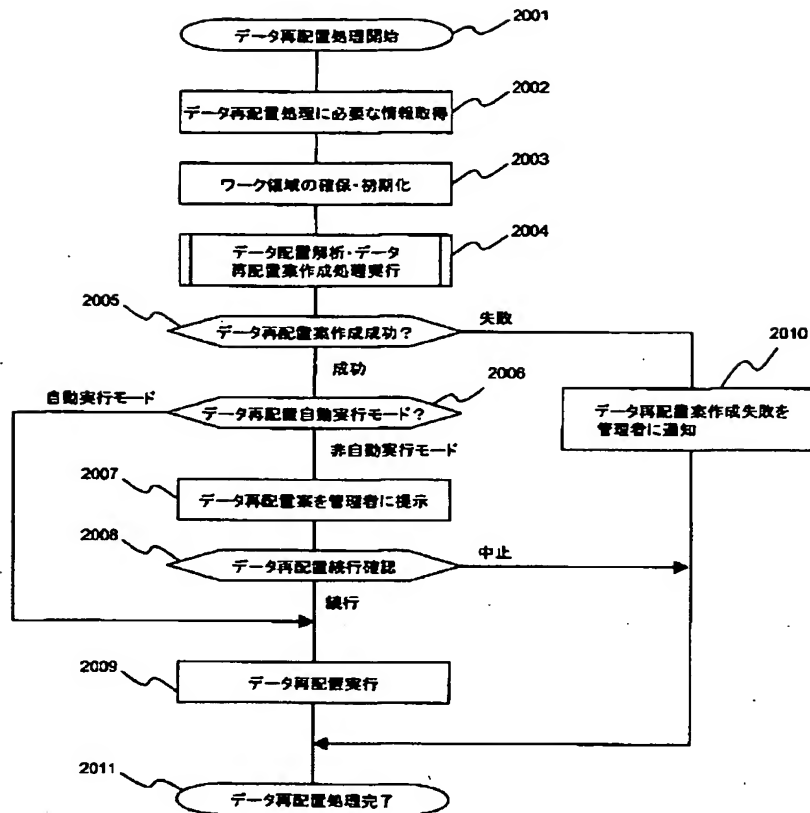


【図16】



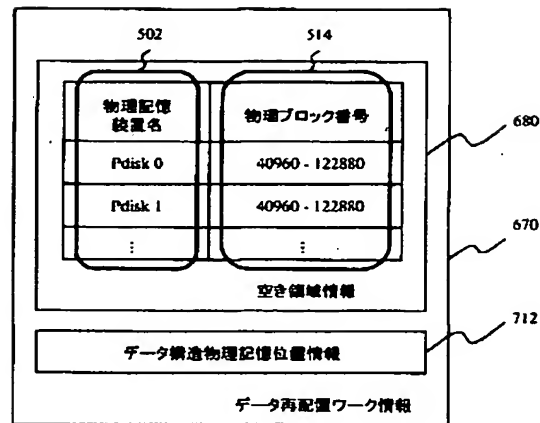
【図 17】

図17



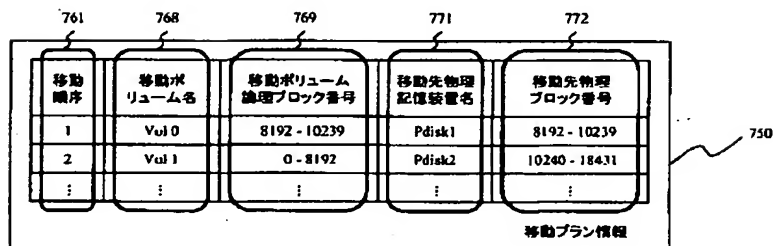
【図 18】

図 18

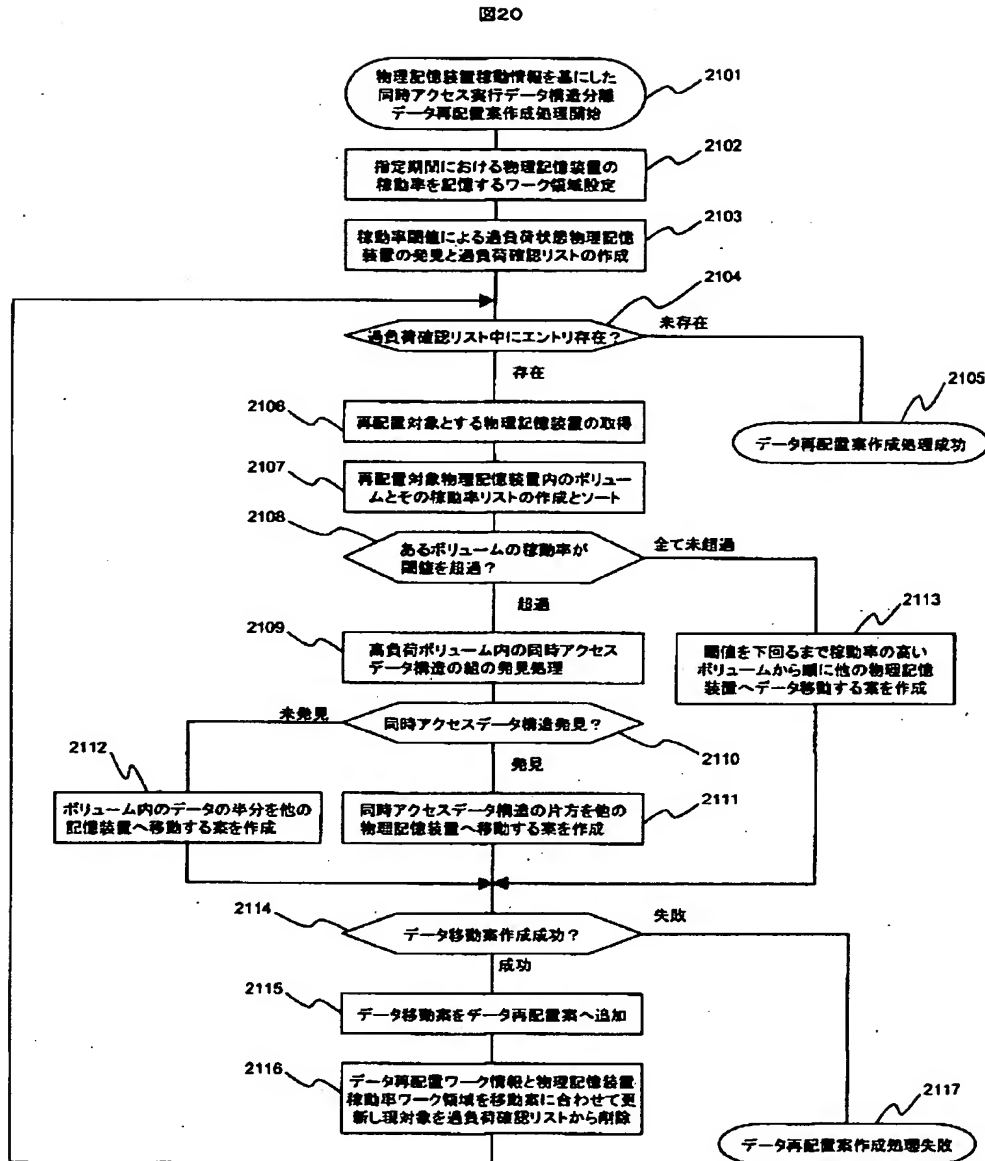


【図 19】

図 19

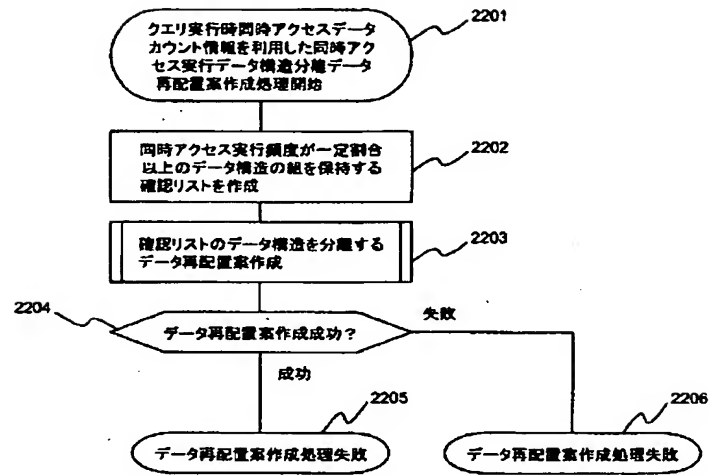


【図 20】



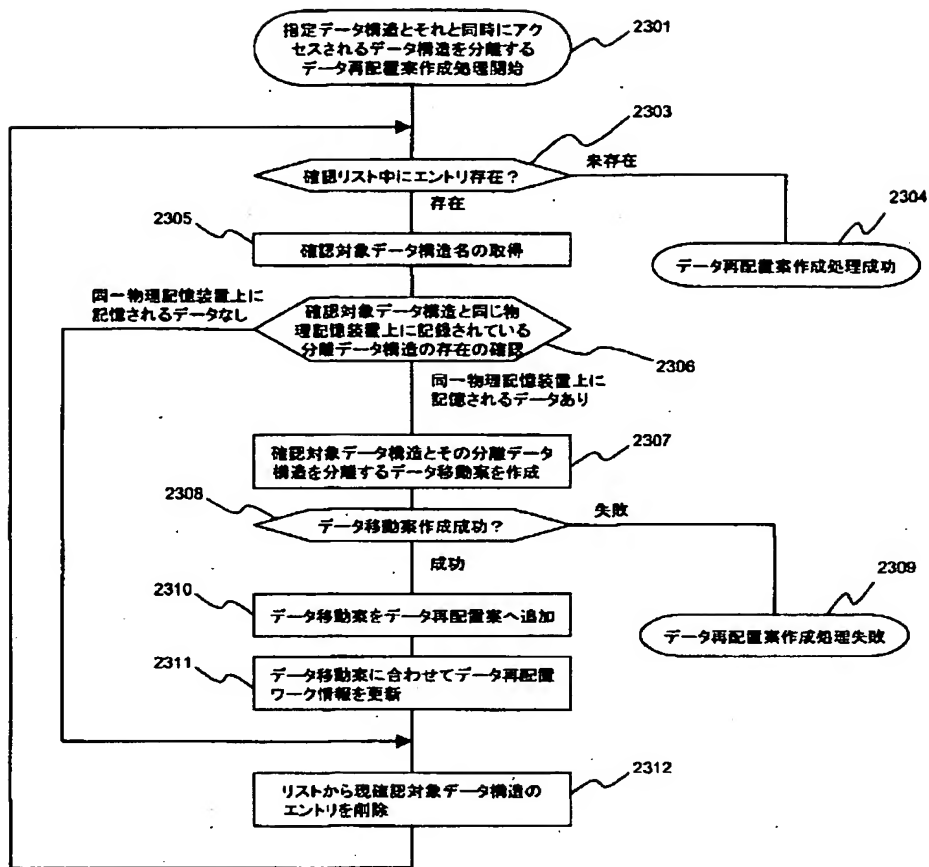
【図 21】

図21



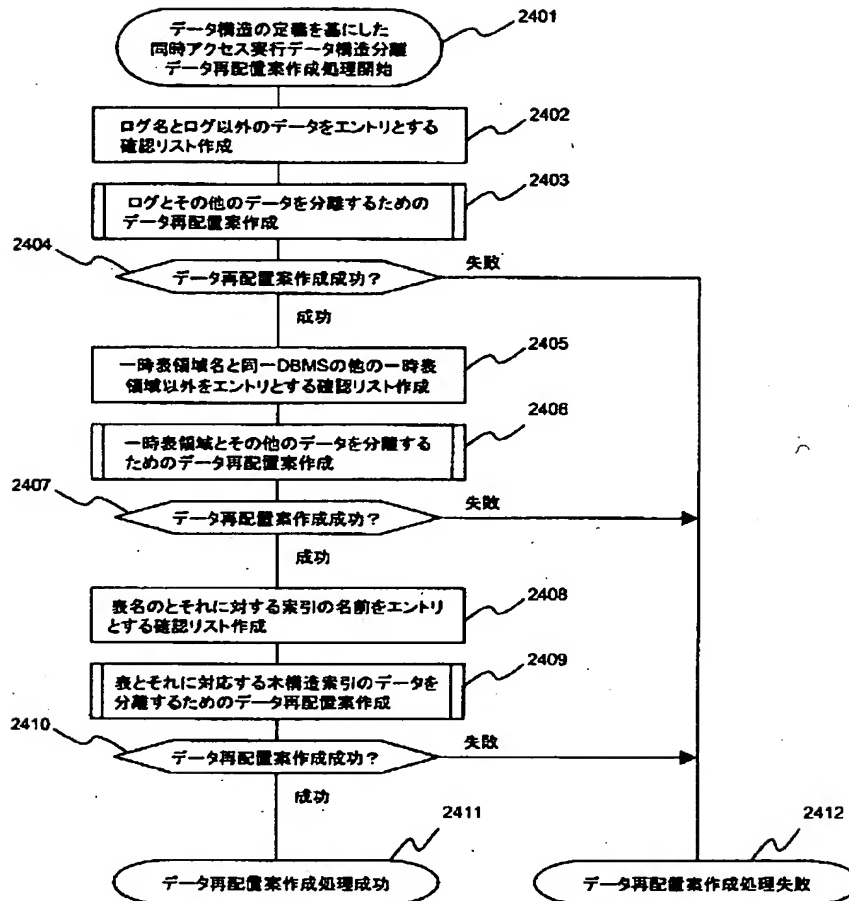
【図 22】

図22

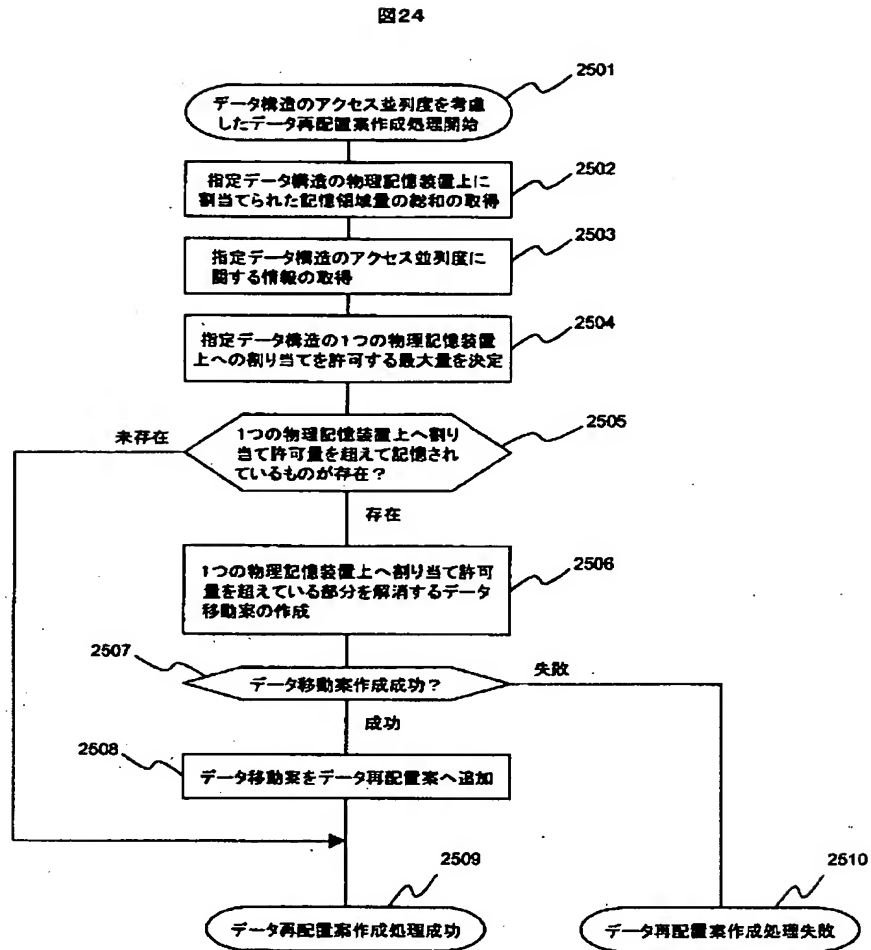


【図 23】

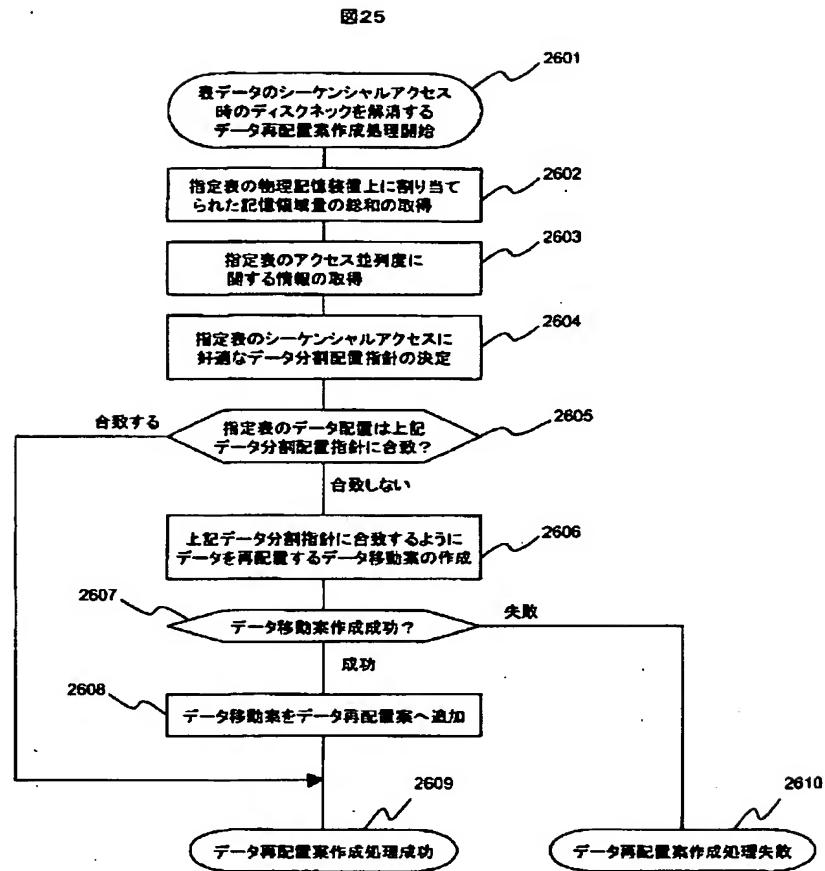
図23



【図 24】

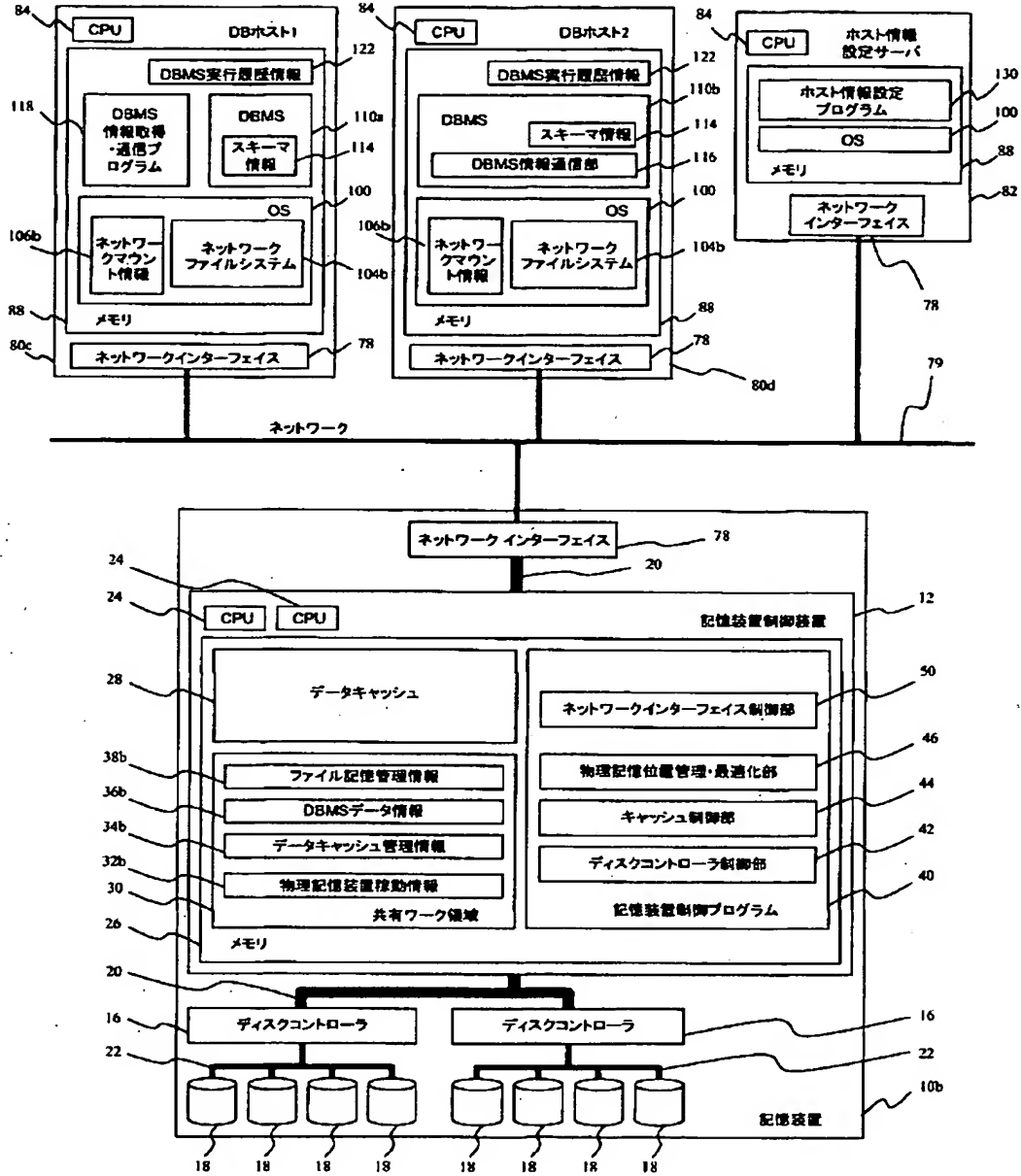


【図 25】



【図26】

図26



【図 27】

図27

記憶装置名	ファイルシステム名	マウントポイント
NAS0	FS0	/NAS0/FS0
NAS0	FS1	/NAS0/FS1
NAS0	FS2	/NAS0/FS2

ネットワークマウント情報

【図 28】

図28

ファイルシステム名	ファイルパス名	ファイルブロック番号	物理記憶装置名	物理ブロック番号
FS 0	/control.dat	0 - 1023	Pdisk 0	4096 - 5119
	/wk/wk1.dat	0 - 2047	Pdisk 1	1024 - 3071
	/wk/wk1.dat	2048 - 4095	Pdisk 0	8192 - 10239
	Empty	-	Pdisk 0	40960 - 122880
FS 1	/data1.dat	0 - 4999	Pdisk 1	10240 - 15239
⋮	⋮	⋮	⋮	⋮

ファイル物理記憶位置情報

ファイルシステム名	ファイルパス名	移動ファイルブロック番号	移動先物理記憶装置名	移動先物理ブロック番号	差分管理情報	コピーポイント
FS 1	/data1.dat	8192 - 10239	Pdisk 1	38912 - 40959	0 1 ... 0	9840
FS 2	/dataA.dat	0 - 8191	Pdisk 1	30720 - 38911	0 0 ... 1	1792
⋮	⋮	⋮	⋮	⋮	⋮	⋮

ファイルデータ移動管理情報

ファイル記憶管理情報

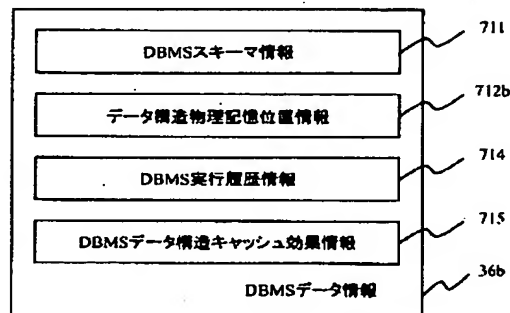
【図 29】

図29

ファイルシステム名		FS0	FS0	FS1	...	1001
物理記憶装置名		Pdisk 0	Pdisk 1	Pdisk 0	...	502
累積稼働時間		23917390	38902849	8012891	...	592
旧累積稼働時間		22787638	38783484	7592039	...	593
稼働率	2000/4/1 12:00 ~ 2000/4/1 12:15	20%	12%	4%	...	594
	2000/4/1 12:15 ~ 2000/4/1 12:30	15%	10%	7%	...	
	2000/4/1 12:30 ~ 2000/4/1 12:45	16%	9%	5%	...	
	
	
前回累積稼働時間取得時刻: 2001/4/12 18:15		物理記憶装置稼働情報				32b

【図 30】

図30



【図 3 1】

図 31

DBMS名	DBMS1	DBMS1	DBMS1	...	DBMS2	...	631
データ構造名	T1	T1	T2	...	Ti	...	361
ファイルシステム名	FS1	FS1	FS1	...	FS2	...	1001
ファイルパス名	/data1.dat	/data1.dat	/data2.dat	...	/datai.dat	...	1002
ファイルブロック番号	0 - 4999	5000 - 9999	0 - 4999	...	0 - 9999	...	1003
物理記憶装置名	Pdisk 1	Pdisk 0	Pdisk 1	...	Pdisk 2	...	514
物理ブロック番号	10240 - 15239	10240 - 15239	15240 - 20239	...	20480 - 30479	...	502
データ構造物理記憶位置情報							712b

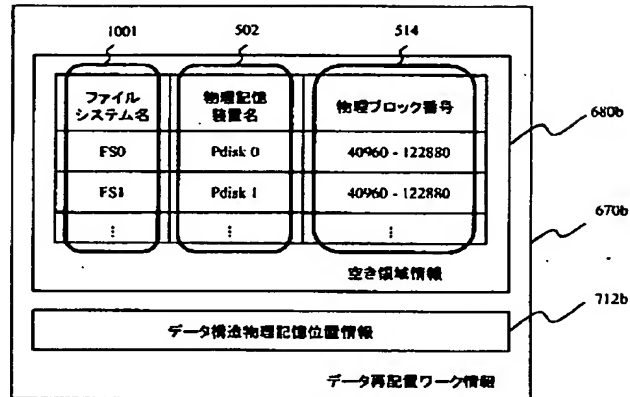
【図 3 2】

図 32

721	1001	1002	1003	722	723
セグメント ID	ファイル システム名	ファイル パス名	ファイル ブロック番号	ステータス 情報	リスト情報
0	FS1	/data1.dat	160 - 175	ダーティ	リスト名 Up Down
1	FS2	/data1.dat	8992 - 9007	ノーマル	第1 342 3922
2	FS1	/data1.dat	512 - 527	ノーマル	第2 409 1523
3	FS0	/control.dat	0 - 15	ライト	第1 4027 1627
...	再利用 - 828
キャッシュセグメント情報					
741	第1LRUリスト情報				742
	MRU セグメントID	LRU セグメントID	セグメント数	MRU セグメントID	LRU セグメントID
	214	672	11776	2890	2332
	第2LRUリスト情報				740
743	MRU セグメントID	LRU セグメントID	セグメント数	再利用LRUリスト情報	
	3	1442	512	データキャッシュ管理情報	
	キャッシュセグメント利用管理情報				34b

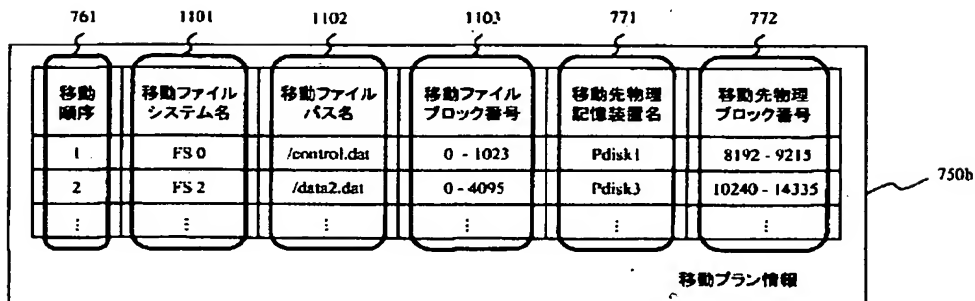
【図 33】

図33



【図 34】

図34



【書類名】 要約書

【要約】

【課題】

データベース管理システム（DBMS）の特性を考慮したデータ配置やキャッシュの制御を行うことにより、記憶装置のデータアクセス性能を向上させる。

【解決手段】

記憶装置は、ネットワーク 7 9 を通して DBMS の静的な構成情報を DBMS 情報取得・通信プログラム、DBMS 情報通信部、ホスト情報設定プログラムを通して取得し、DBMS データ情報としてメモリ内に記憶する。記憶装置制御プログラム内の物理記憶位置管理・最適化実行部は DBMS データ情報を利用してデータ再配置を実行し、キャッシュ制御部は DBMS データ情報を加味したデータキャッシュ制御を行う。

【選択図】 図 1

認定・付加情報

特許出願の番号	特願2001-345525
受付番号	50101661496
書類名	特許願
担当官	第七担当上席 0096
作成日	平成13年11月13日

<認定情報・付加情報>

【提出日】	平成13年11月12日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日 1990年 8月31日
[変更理由] 新規登録
住 所 東京都千代田区神田駿河台4丁目6番地
氏 名 株式会社日立製作所